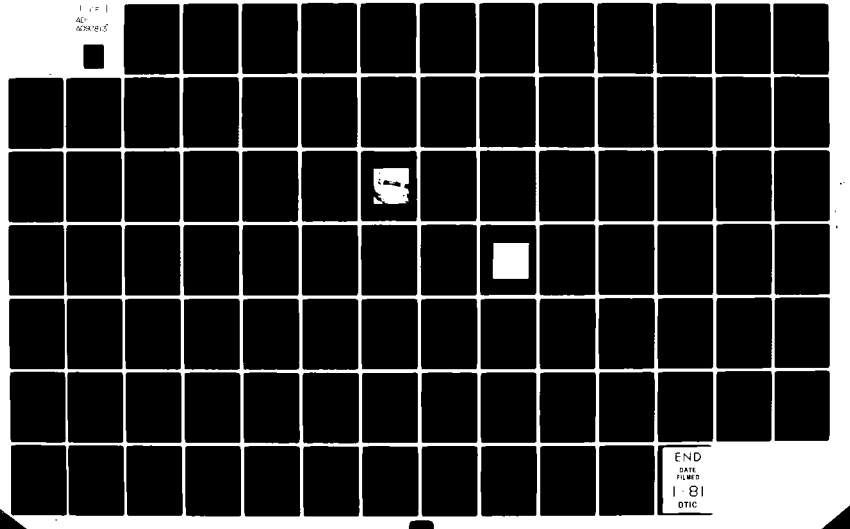
AD-A092 813 E-SYSTEMS INC DALLAS TEX GARLAND DIV F/G 9/2
STUDY OF DIGITAL MATCHING OF DISSIMILAR IMAGES.(U)
OCT 80 D S RHINES DAAK70-79-C-0235
UNCLASSIFIED ETL-0244 NL



ETL-0244

LEVEL II

AD A092813

FINAL REPORT, STUDY OF DIGITAL MATCHING OF DISSIMILAR IMAGES

Don S. Rhines
E-SYSTEMS, Inc.
Garland Division
P.O. Box 226118
Dallas, Texas 75266

31 October 1980

Approved for public release; distribution unlimited

Prepared for

U.S. Army Engineer Topographic Laboratories
Fort Belvoir, Virginia 22060

Destroy this report when no longer needed.
Do not return it to the originator.

The findings in this report are not to be construed as an official Department of the Army position unless so designated by other authorized documents.

The citation in this report of trade names of commercially available products does not constitute official endorsement or approval of the use of such products.

UNCLASSIFIED
SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER ETL-0244	2. GOVT ACCESSION NO. AD-A092 823	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle) Final Report, Study of Digital Matching of Dissimilar Images.		5. TYPE OF REPORT & PERIOD COVERED Contract Report 6. PERFORMING ORG. REPORT NUMBER
7. AUTHOR(s) Don S. Rhines		8. CONTRACT OR GRANT NUMBER(s) DAAK70-79-C-0235
9. PERFORMING ORGANIZATION NAME AND ADDRESS E-SYSTEMS, Inc., Garland Division P.O. Box 226118 Dallas, Texas 75266		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS 12 93
11. CONTROLLING OFFICE NAME AND ADDRESS U.S. Army Engineer Topographic Laboratories, Fort Belvoir, Virginia 22060		12. REPORT DATE 31 October 1980 13. NUMBER OF PAGES 98
14. MONITORING AGENCY NAME & ADDRESS(if different from Controlling Office)		15. SECURITY CLASS. (of this report) UNCLASSIFIED 15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

Approved for public release; distribution unlimited

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

DTIC ELECTE DEC 8 1980 S A

18. SUPPLEMENTARY NOTES

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

Matching, Dissimilar Images, Physical Commonalities, Feature Extraction, Similarity Measures

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

This final report presents the results of a study conducted for the U.S. Army Engineering Topographic Laboratories on the digital matching of dissimilar images. This report develops a practical approach for the digital determination of corresponding points on dissimilar images. This approach could be used to register a large number of points automatically

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE

UNCLASSIFIED 408959
SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

in a reasonably short period of time. The algorithms required are presented in a manner that can be coded in FORTRAN IV and tested on the DIAL facility at USAETL.

PREFACE

This report is generated under Contract DAAK70-79-C-0235 for the U.S. Army Engineer Topographic Laboratories, Fort Belvoir, Virginia 22060 by E-SYSTEMS, Inc., Garland Division, Dallas, Texas and submitted as ETL-0244. The Contract Officer's Representative was Mr. Tom Blackburn.

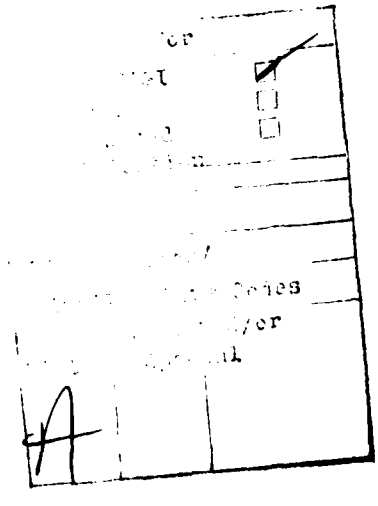


TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF FIGURES (Continued)

1.0 Introduction

This final report presents the results of a study conducted for the U. S. Army Engineering Topographic Laboratories on the digital matching of dissimilar images. This report develops a practical approach for the digital determination of corresponding points on dissimilar images. This approach could be used to register a large number of points automatically in a reasonably short period of time. The algorithms required are presented in a manner that can be coded in FORTRAN IV and tested on the DIAL facility at USAETL.

For similar images, conventional correlation techniques can be used to digitally match images. With the measurement of many corresponding points, contour information can be developed using this approach. However, for the technique to work with any degree of reliability, requires that special optical cameras be used and that the taking conditions be precisely controlled. The photographs must be taken at the same time of day from just the right perspectives.

This report deals with the study of techniques that would permit digital matching between a much broader class of images. It is desired to be able to match images from different detector sources. Optical, infrared, and radar detectors are examples of the broad source of images it is desired to be able to match. Similarly, it is desired to be able to match images from a single sensor when the images are taken from very different perspectives or under different illumination conditions. The most difficult class of images to match are radar images. This is because of the coherent nature of the imaging techniques. Several specific suggestions are included in this report to help deal with radar images.

There are two different approaches that can be taken to develop techniques to match dissimilar images. One could attempt to use a pattern recognition approach or a texture matching approach. In a pattern recognition approach, items in the scene are identified through feature extraction. In a texture matching approach, it is assumed that between matching areas of the images there will remain matching textural components. With the texture matching approach, the difficult task of item identifiction is unnecessary. In this study the texture matching approach to matching dissimilar images has been used.

In most situations there will exist some information about the taking conditions used to generate the images of interest. This information could include the approximate position of the camera including altitude and the aiming direction used. To reduce the dimensions of the matching problem, it is strongly suggested that whatever a priori information is available be used initially in the matching process to generate geometric corrections. In Section 2 a general technique is presented to develop geometric corrections. This technique can be applied to optical or radar images, and can include the use of contour information if available. After the application of these corrections, the dominant misregistration components will be translational. Scale and rotational differences will normally be small. Both global and local translational differences will exist. The global or overall translational differences come from the residual uncertainties in aim point determination. Local translational differences are due to the combined effects of the perspectives used and uncorrected terrain effects. It is generally assumed in developing matching algorithms that geometric corrections of this type have been applied.

There are two extreme applications to which the matching of dissimilar images could be used. On the one hand, it could be desired to match thousands of points for the development of a contour data base. On the other hand, it could be desired to measure a few match points from each image in order to approximately register the images. After approximate registration, the global translational differences between the two images have been removed, but the local translational differences would still remain. The approximate registration of images might be required in a map updating application. Also, approximate registration of the images could be used as the first step of a contour measuring process. The texture matching approach that has been developed in this report could be used in either of these extreme applications.

In addition to the texture matching approach, another approach is developed for use in the approximate registration of images. This approach is described in Section 3, and is a semi-automatic approach which makes use of the human ability to find match points between differing images. The approach provides computer and special-purpose hardware assistance to the operator who is required to measure matching points from each of the images. This approach is referred to as computer assisted manual image registration (CAMIR). This system could be readily constructed given the current state-of-the-art of digital hardware.

The texture matching approach has the advantage of being completely automatic. The basic concept is to replace the actual sample values, which are very dependent upon the collection system, with derived values which are independent of the collector system. With system independence and with the application of gross

geometrical corrections, image pairs can be registered by conventional correlation techniques. The key to this approach is the development of specific conversion algorithms for the pixels of each system of interest, so that the resultant pixel values are effectively system independent. Although the general procedure can be readily described and implemented, and although there is considerable experience with specific transformations, the actual pixel replacment algorithms will have to be empirically tailored for specific collection systems.

In Section 4 the texture matching concept is demonstrated. A single aerial image is used. It is found that when this image is modified with the use of an edge enhancement operation, match points can no longer be obtained using a conventional correlator. The pixels from both the original and modified images are replaced with a measure of local texture. The texture measure used, referred to as the local busyness measure, uses the local AC energy present in the neighborhood of the replaced pixel. It is then found that conventional correlation techniques can be used to accurately match the images. Factors affecting match point resolution are further discussed in this section. It is found that it is not necessary to replace each and every pixel with a texture measure in order to maintain match point resolution. The frequency with which pixel intensity values are replaced is referred to as the pixel replacement resolution.

The texture matching approach is further developed in Section 5. Derivation of local texture measures through the use of local spectral analysis is discussed. The local spectral analysis is performed with the use of the two dimensional discrete Fourier transform. This is a powerful approach which can be used to

generate a broad class of texture measures. The texture matching approach is formally defined, and the technique is generalized to permit the use of multiple texture measures. Finally, a variety of special problems which can interfere with texture matching are discussed. Methods are discussed for dealing with data noise characteristics, different detector spatial acceptances, drastically different intensity distributions, and coherency effects found in dealing with imaging radars.

2.0 GEOMETRIC CORRECTIONS

In our approach to the problem of matching a large number of points between the dissimilar images, the first step is to make all possible geometric corrections based upon a priori knowledge. In this section geometric correction fundamentals are outlined.

The first step is to derive a relationship between positions in the field-of-view of the taking system and the corresponding points onto which they are projected by the taking system. Let positions in the field-of-view of the system or in object space be denoted by the coordinates x, y, and z. Let positions on the projection surface be denoted by the coordinates α and β. Then the desired relationships are of the form:

$$\alpha = f_{\alpha}(x,y,z) \tag{2-1}$$

$$\beta = f_{\beta}(x,y,z) \tag{2-2}$$

To derive these relationships, the taking parameters which would need to be known would include the position and altitude of the system, the aiming direction of the system, as well as other specifics of the projection process used by the system. In some cases a simple analytic expression can be fairly easily derived. An example of this would be a snapshot camera with no lens distortions present. In other cases much more elaborate procedures

would be necessary to derive the appropriate relationship. An example of this case would be Landsat satellite imagery. In this case the satellite flight path and motion used during the exposure must be taken into account.

The next step is to determine a grid of points on a surface in object space. The original scene is distorted due to the taking perspective and characteristics of the taking system. We would like to determine the appropriate pixel intensities in an undistorted space, and the grid of points required gives the positions at which we would like to know the pixel intensities. Let position on this object surface be denoted by the coordinates r and s. Since we are developing information for use in map making, the coordinates r and s will usually correspond with appropriate map projection coordinates. An example would be the position on the surface of the earth as specified by longitude and latitude. If detailed contour information is available, then the surface to be used can correspond with the actual surface of the earth. If detailed contour information is not known, then an approximation to the earth's surface at the appropriate average altitude can be used. If the area covered is small (less than a square mile), then a flat plane would be adequate for this surface. For larger area coverage, then a sphere or oblate spheroid surface would be an appropriate approximation. For the choice of grid points made, then for each

grid point its position in object space must be determined. In functional form, the required relationship is given by:

$$x = f_x(r,s) \tag{2-3}$$

$$y = f_y(r,s) \tag{2-4}$$

$$z = f_z(r,s) \tag{2-5}$$

The next step is to determine for each grid point the corresponding position in the projection plane. Substituting Equations (2-3), (2-4), and (2-5) into (2-1) and (2-2), we obtain the desired relationships:

$$\alpha = f'_\alpha(r,s) \tag{2-6}$$

$$\beta = f'_\beta(r,s) \tag{2-7}$$

For many cases it should be possible to derive this relationship directly. This can be done as long as the object surface can be described by some simple functional form. The one clear exception is the case in which contour information is being used. In this case, the altitude information would be contained in a grid of altitude values. In this case it would not be reasonable

to derive explicitly relationships (2-6) and (2-7). Instead, a two step procedure would be required. First, for each grid point, its position in object space must be determined by interpolation from the altitude data base. Next, the corresponding position on the projection surface can be determined using Equations (2-1) and (2-2).

In the final step, the intensity values associated with each of the grid points which have been mapped onto the projection plane are determined in an interpolation process. The interpolator requires the position of the grid point with respect to the pixel locations and the value of the intensity levels associated with the pixel locations in the immediate neighborhood of the projected grid point. In many applications a 2x2 point linear interpolator should give adequate results. Performance can be improved by going to a higher order interpolator using specially developed coefficients. At E-Systems, we have extensive experience in developing coefficients for higher order interpolators should high performance be required.

When hardware implementations of geometric corrections are considered, it becomes clear that including contour information adds considerably to the complexity of the problem. If contour information is to be used, the process probably should be divided up into a two-stage process. Initially, the geometric corrections should be generated without the use of the contour

information. The image can then be registered with respect to a reference image. Using this process the knowledge of the taking parameters can be refined. Without this initial step, knowledge of the taking parameters would not normally be sufficiently accurate to warrant the use of contour information. In the second stage, the full geometric corrections with the use of the contour information can be generated.

There is one other special problem that can arise with the use of contour information. In general, if the surface is bumpy or mountainous and the look angle well off nadir towards the horizon, then there will be regions on the object surface which will be hidden from the detector. For geometrically corrected optical imagery, these regions should be flagged in some unambiguous manner.

Consideration of the addition of contour information should depend on the class of images to be used in making maps. In many map updating operations, contour information should be frequently available and be useful. The more mountainous the terrain, the larger will be the distortions that can be generated. In general, the more an optical detector looks off nadir toward the horizon, the more severe will be these distortions.

3.0 COMPUTER ASSISTED MANUAL IMAGE REGISTRATION

To solve the problem of matching dissimilar images, it is suggested that the first step is to bring images into approximate registration. In this section we describe a technique that can be used to obtain approximate registration. This technique makes use of high-speed digital approaches to applying geometric corrections, and makes use of the exceptional ability of a human operator to quickly solve complex pattern recognition problems. This approach is referred to as computer assisted manual image registration (CAMIR). The state-of-the-art of digital hardware has advanced to such a point that this system could be readily constructed. It is felt that this approach offers the most practical immediate solution to the problem of quickly bringing into approximate registration dissimilar images.

First, we will describe the operation and performance capabilities of the system that is envisioned. The system will require both the use of a general-purpose computer and the use of high-speed special-purpose hardware. The particular approach required to correct geometric distortions will be described. This unique approach permits the real-time implementation of the geometric corrections for any type of detector using high-speed digital hardware. Details of the hardware are beyond the scope of this study, but cursory analysis indicates that the hardware requirements are well within the current state-of-the-art.

To view each of the two dissimilar images, the operator will view a CRT display. The CRT would normally display just a small portion of the image. The operator would have complete freedom to explore and examine a given image. This freedom would include the ability to translate, to rotate, and to both magnify or demagnify the sub-image being viewed. Updating of the sub-image would be performed at a sufficiently high rate to assure smooth presentation of manipulations to the display. Geometric corrections based on a priori knowledge are normally applied before the image is projected onto the CRT.

To approximately register the two images, the operator would select three or more corresponding points. For best performance these points should be selected as much as possible at the extremes of the images, as widely separated as possible. In order to make the measurements, the operator would bring a cursor projected on top of the CRT display into coincidence with the point to be measured.

After three corresponding points in each image have been measured, the geometric corrections would be refined to bring the two images into approximate registration. The three points which have been measured define a plane. All points which lie in this plane will be now exactly registered. Points lying above or below this plane would be shifted with respect to one another from one image to the next. The amount of shift would depend on how far the points are from the registration plane and on the difference in the taking perspectives. For many pairs of

images the remaining shifts present over most of the field-of-view would be quite small. This fact can now be used in developing further techniques to register many points between the images.

If it were desirable, it is possible at this point to develop a system which would simultaneously project the two images for stereo viewing. As long as the taking perspectives are appropriate, the two images could be from completely different taking systems. For stereo viewing of any portion of the image, no further manually applied corrections would be necessary.

There are several other highly desirable features which could be included under the real-time control of the operator. One of these is the ability to adjust the correction being applied to account for the spatial frequency response found in optical systems. This correction is known as modulation transfer function compensation (MTFC), and is performed by the convolution of an MTFC matrix with the scene data. Under a real-time implementation, the MTFC could be modified at the frame rate. This allows real-time control of image sharpness and noise. Another highly desirable feature would be the ability to modify the transformation of the scene intensities under the real-time control of the operator. This transformation, known as dynamic range adjustment (DRA), can be implemented by the use of a look-up table to map intensity levels in any appropriate manner. An example of its use would be to bring out details in a shadow area.

The key to the implementation of this system is in the approach taken to handling the geometric distortions. The solution to the problem requires both the use of a computer and of special-purpose real-time hardware. To handle a variety of taking systems, the problem must be cast into a uniform form for use by the special-purpose hardware by using software within the computer.

The first step is to generate the mapping from the object surface (r,s) to the projection surface (α,β) in polynominal form. In other words, **Equations** (2-6) and (2-7) must be found in the form:

$$\alpha = f'_{\alpha}(r,s) = a_1+a_2r+a_3s+a_4r^2+a_5rs+a_6s^2+... \qquad (3\text{-}1)$$

$$\beta = f'_{\beta}(r,s) = b_1+b_2r+b_3s+b_4r^2+b_5rs+b_6s^2+... \qquad (3\text{-}2)$$

These polynomial coefficients must be generated by the computer software and be passed on to the special-purpose hardware previous to any viewing of the images. In this manner corrections for any type of detector can be cast into a single form which can be handled by special-purpose hardware.

In order to determine polynomial coefficients, a grid of points must be mapped from the object surface to the projection surface using the appropriate transformation for the given detector derived from the initial taking parameters. The number

of grid points required depends upon the degree of the polynomial. In general, the number of points used must be greater than the number of coefficients required for one of the polynomials. For example, for a seventh order polynomial requiring 36 coefficients, a 9x9 grid of 81 points should be sufficient. Once the grid points have been mapped onto the projection surface, the polynomial coefficients are solved for by the least squares fitting which minimizes the sum:

$$\sum_{i,j} \left\{ \left(\alpha^{m}_{i,j} - \alpha^{f}_{i,j} \right)^2 + \left(\beta^{m}_{i,j} - \beta^{f}_{i,j} \right)^2 \right\} \qquad (3\text{-}3)$$

The degree of the polynomials required can only be determined by analysis of the taking parameters for the systems of interest.

These geometric correction polynomials permit correction of distortions over the entire extent of the image. Usually the CRT will display only a small portion of this, and in this local area it is possible to describe the distortions by a much

smaller polynomial. It is suggested that for viewing purposes only that the distortions be described by a second order polynomial. An appropriate CRT resolution for displaying local regions of the image might be 512x512 pixels. We will use this resolution for example purposes. The derivation of the appropriate second order polynomial from the higher order polynomial must be evaluated in special-purpose hardware. It is this polynomial evaluation hardware which permits the translation, rotation, and magnification of the image by the operator at frame rates. To derive the second order polynomials, a change of coordinates is performed. We explain the procedure in terms of a simple translation example. It is desired to display the sub-image centered at the point (r_0,s_0) on the object surface. A transformation of coordinates is performed which maps this point into the point (0,0). This coordinate transformation is given by:

$$r' = r-r_0 \tag{3-4}$$

$$s' = s-s_0 \tag{3-5}$$

The correct second order polynomials are obtained by substitution of this change of coordinates into the polynomial correction Equations (3-1) and (3-2). The terms out to second order in r' and s' are gathered together to give the second order polynomials. The CRT then displays the interpolated intensity values

at locations corresponding to r´ going from -256 to 256 and s´ going from -256 to 256. The ability to rotate and magnify or demagnify the sub-image being projected is obtained in the same manner by the use of the appropriate change of coordinate equations.

After the second order polynomials are found, the next step determines where each of the 512x512 points to be displayed fall in projection space (the space in which the original measurements are available). This is performed by a polynomial solver special-purpose hardware which evaluates the second order polynomial for the 512x512 points to be displayed. Because the transformation is now only of second order, it is possible to design the hardware to find each succeeding point by using only additions. Reduction of the polynomial corrections locally to second order results in reducing the hardware requirements substantially at this point.

The final special-purpose hardware required is a two-dimensional interpolator. The inputs to this interpolator include the original pixel measurements and the position in projection space of each of the 512x512 pixels to be displayed. From the position of each pixel to be displayed and from the original intensity measurements in the immediate vicinity, the appropriate intensity values to be displayed are derived.

Once the three corresponding points have been measured by the operator, the images can now be brought together into registration. This requires an updating of the distortion polynomials. These corrections must be calculated in the computer. The technique uses a linear transformation of both the images to bring them together. The linear transformation is of the form:

$$r' = c_1 r + c_2 s + c_3 \qquad (3\text{-}6)$$

$$s' = c_4 r + c_5 s + c_6 \qquad (3\text{-}7)$$

The technique used to derive the coefficients finds the optimum amount to move each of the images, given the errors on the initial taking parameters. Updating the polynomial corrections requires a substitution or change of variables operation similar to the technique used to provide for the translation of the sub-image being viewed.

To be practical, computer aided image matching must be capable of processing huge amounts of image data. A typical frame might come from an original photographic image which is 10X10 inches square. The sampling might be 2000 points per inch. The image data for this case is an array of 20,000 by 20,000 samples. The display and processor must be capable of handling at least one pair of these images and should be capable of providing for many pairs.

A CRT display is only 512x512 pixels or 1024x1024. Larger displays could be developed, perhaps, but are not really needed. In fact, they are not even desirable. If we had a CRT with 20,000 by 20,000 pixels resolution, we would need a microscope for most image examination since we could not resolve the image detail by eye alone. Another way to achieve the required magnification on such a super CRT, so that the eye can perceive detail, would be to digitally magnify the image and display only a small image portion around the point of interest. This implies that many CRT pixels are used to display each original sample. For example, for 10X magnification, the full screen would be used to display an area only 2000 by 2000 samples. One hundred CRT pixels would be used to display each original sample. Of course, the effect of magnification by digital interpolation is to smooth the display and provide optimum resolution. So each original sample is not just replicated, but on the average,

one hundred pixels are used for each original sample.

This is just like examination of film with a microscope. The eye can see the whole frame, but cannot resolve all the detail. The field-of-view of the microscope covers only a small portion of the film image area and provides details.

A super high resolution CRT is not needed. When the entire frame is displayed, the eye can not resolve fine detail anyway. A 512x512 or 1024x1024 CRT can provide all of the resolution, detail and field-of-view that a human operator can interpret.

To make a normal CRT applicable to image matching, the original samples must be processed to provide displays ranging from full frame to image areas small enough to provide full resolution and detail. Further, this processing must be rapid enough to follow the operator's manipulations. Taking clues from the use of light tables and microscopes, we note that the human operator examines the whole frame or large portions of the frame, centers a microscope over an area of interest and observes this area in greater detail. Often the operator alternates between large and small scale image examination, varying magnification and field-of-view and slewing about over an image. Actually, the field-of-view of a zoom microscope varies only in respect to the film image. The virtual image presented to the eye is of fixed extent, just as the screen of a CRT is of fixed size.

What is needed for practical, computer aided manual image registration is a digital processor, which can handle pairs of 20,000 by 20,000 sample images and provide arbitrary slew and zoom for a CRT by full image alteration during the 1/30th second framing period. Such a device would provide all of the capability of a light table and zoom microscope coupled with real-time image geometry manipulation.

This digital processor and display for image matching is well within the current state-of-the-art. E-Systems could build this image manipulator with only engineering effort - no research or development effort would be needed.

4.0 TEXTURE MATCHING CONCEPT DEMONSTRATION

In this section the concept of texture matching is demonstrated with the use of digital data from a single aerial image. This example illustrates texture matching as an extension of conventional correlation matching in which each pixel value is first replaced by some measure of the texture in its immediate neighborhood.

In the first subsection, a sample aerial image is used to investigate a conventional correlation technique using the normalized cross correlation product. The object here is to derive a second image which no longer correlates with the original image. This second image is derived by modifying the original image. Several modifications are investigated including the addition of noise, low pass filtering, a nonlinear intensity transformation, and an edge enhancement operation. The edge enhancement operation is found to produce an image which no longer correlates with the original image.

In the next subsection, a simple texture matching approach of finding match points is investigated. The original image and the image modified by the edge enhancement operation are used. Each pixel in both images is replaced by a measure of the local activity in a four by four neighborhood of each pixel. After this replacement, the normalized cross correlation approach is found to produce reasonably accurate match points. Next, factors affecting the accuracy of the match points are investigated. Accuracy is investigated for procedures which replace every other pixel in every other line and every fourth pixel in every fourth line by the same measure of local activity. Using these alternative procedures, roughly the same accuracy is obtained.

4.1 Conventional Correlation Experiments

Figure 4-1 shows the aerial image that has been used in Section 4 to demonstrate the texture matching approach to finding image match points. The digital representation of the image has been obtained by scanning the original picture with a micro-densitometer. The digital representation of the image consists of approximately 1000 by 1000 pixels. Each pixel is represented by a ten bit reflectance value measured by a densitometer. The digital image has been reconstructed to form the picture shown in Figure 4-1 through the use of a laser scanning device.

In this subsection we would like to find modifications of the original image which prevent the use of correlation techniques to find match points between the original and the modified image. Hopefully, these modifications will be representative of the type of problems present in trying to match dissimilar images of the type we would like to be able to match. Then, in the next subsection, the texture matching extension of the correlation match point technique can be demonstrated on these two images.

In this subsection the normalized cross correlation approach is used to determine match points between the original image and the modified image derived from the original. This is a powerful approach which is very effective for determining match points between similar images. For convenience, the normalized cross correlation is defined here for the one dimensional case. The generalization to two dimensions is straight forward. Let x_i be the reference window which we would like to match with some area of a search area given by the data y_i, i.e.

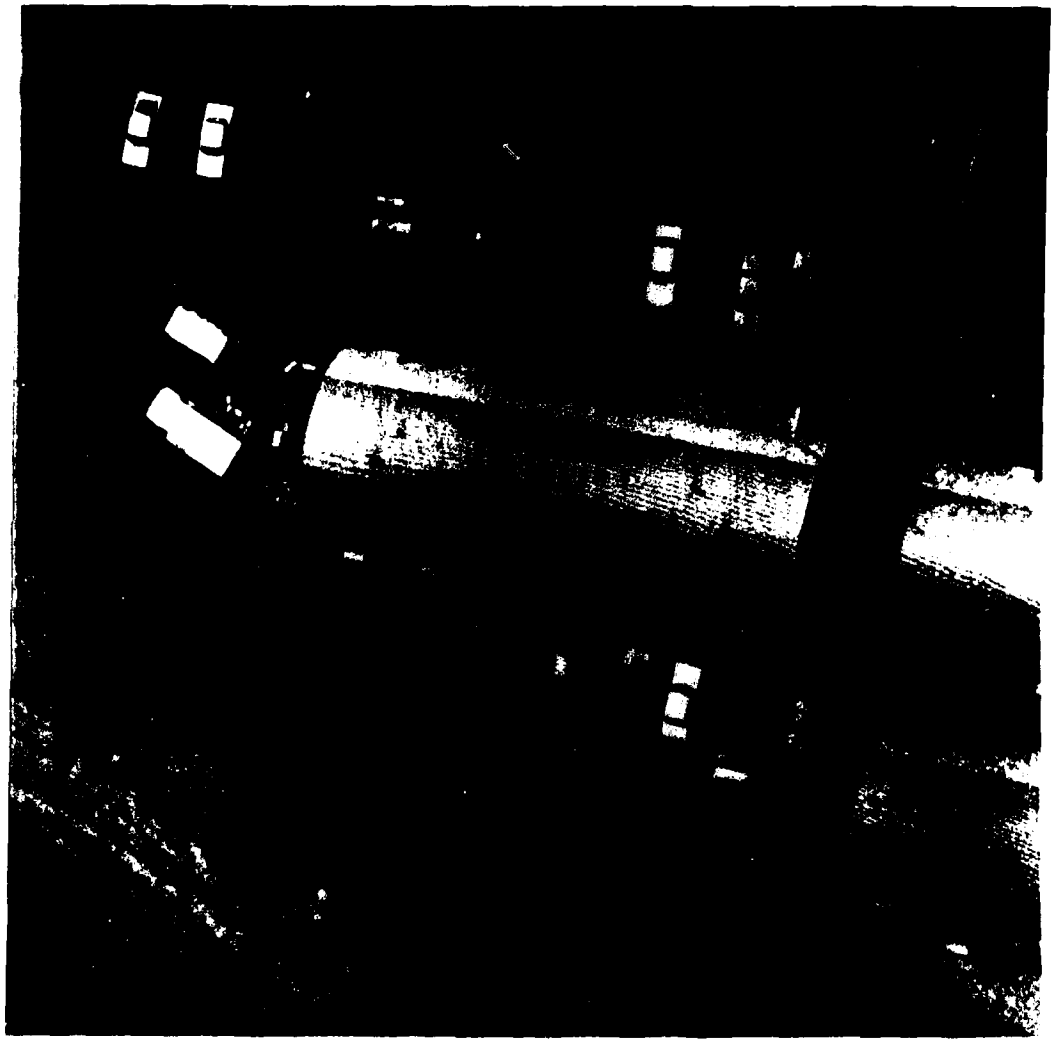


FIGURE 4—1.

SAMPLE AERIAL IMAGE

reference window: x_i, $i = 1, 2, \ldots, N$

search area: y_i, $i = 1, 2, \ldots, M$, $M>N$.

The normalized cross correlation ρ_j, $j = 1, 2, \ldots, M-N+1$, is defined by

$$\rho_j = \frac{\frac{1}{N} \sum_{i=1}^{N} (\hat{X}_i \cdot \hat{Y}_{i+j})}{\sigma^X \cdot \sigma^Y_j} \tag{4-1}$$

with $$\hat{X}_i = \hat{X}_i - \bar{X}, \quad \bar{Y}_j = \frac{1}{N} \sum_{i=1}^{N} Y_{i+j} \tag{4-2}$$

and $$\sigma^X = \sqrt{\frac{1}{N} \sum_{i=1}^{N} (\hat{X}_i - \bar{X})^2} \tag{4-4}$$

$$\sigma^Y = \sqrt{\frac{1}{N} \sum_{i=1}^{N} (\hat{Y}_{i+j} - \bar{Y}_j)^2} \tag{4-5}$$

This normalized cross correlation is bounded between plus and minus one. High correlation corresponds with numbers near one, and match points are determined by searching for this correlation peak. If it is necessary to search large areas, one problem which can arise is multiple solutions. In general, if two data sets differ in their intensity distributions by only a multiplication or a translation factor, i.e.

$$y_i = Ax_i + B \tag{4-6}$$

then their cross correlation product will yield exactly one. This can be demonstrated by substitution into the above defining equations. This is an important property which permits the determination of

match points for a much broader class of similar images than would otherwise be true. This property is due to the normalizing procedure used.

4.1.1 No Modification Example

Before attempting to modify the sample aerial image, nominal correlation results are obtained with no modification introduced. The general procedure followed in each case is identical. A reference window size of 32 by 32 pixels is used. A search area of 48 by 48 pixels centered over the reference window is used. This yields a 17 by 17 element correlation space. With zero error in the match point determination process, the correlation peak occurs at the center element of the correlation space (the ninth element of the ninth row). In the general case, the reference window is taken from the modified version of the aerial image and the search is taken from the unmodified version of the image.

Ten areas have been chosen at random from the image for the correlation examples. For each case (i.e. for the case of no modification and for each of the four possible modification cases) the identical ten areas are used. In other words, the positions of the reference windows and the search areas are unchanged.

When correlation is performed without the introduction of any modification, well defined correlation peaks are obtained for each of the ten examples. Figure 4-2 shows the correlation results for two of the ten areas. These two areas are referred to as Area A and Area B. The correlation numbers in the figure have been scaled by multiplying by a thousand. Thus a perfect correlation of one appears in the figure as 1000. For ease of reading the center element of the correlation results has been boxed out and the correlation peak has been circled.

AREA A

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
1	91	12	3	-47	-16	-15	-24	-28	-12	-33	24	-16	-31	-27	-34	32	-20
2	-56	-30	-14	61	-9	-19	4	-12	27	-28	-28	-45	-26	-22	-2	0	23
3	-9	18	18	4	60	32	27	-9	2	20	29	17	34	17	-20	4	-9
4	96	41	11	45	26	40	9	0	20	81	35	-36	-63	5	40	16	-9
5	-29	34	50	6	-23	-19	20	-5	57	-23	88	58	54	61	38	108	40
6	14	65	-35	21	-48	17	37	-12	-20	41	18	5	48	44	11	52	32
7	13	63	66	21	39	21	47	63	56	31	40	71	93	51	68	48	79
8	-7	21	38	22	41	60	70	99	101	97	50	16	21	45	34	88	53
9	60	47	33	6	-9	8	141	250	1000	251	119	3	-18	-7	32	77	53
10	47	81	13	52	31	12	54	107	100	107	59	49	33	28	46	52	-11
11	33	0	32	45	80	81	39	58	55	62	33	-10	33	16	64	57	28
12	58	55	5	-1	26	-12	-8	23	-23	7	33	7	-44	1	-20	69	8
13	47	102	0	3	23	27	94	-27	41	-13	2	-27	-1	49	12	23	-43
14	16	13	34	-13	-68	-16	4	56	-2	6	2	-14	-12	-15	-27	-4	-8
15	13	11	-16	-30	-10	-18	7	-5	4	-16	89	3	32	-38	-23	-31	-30
16	30	26	0	-24	-39	-42	14	5	32	-2	-19	-31	-34	20	-22	-1	-38
17	-17	25	-25	-35	-56	-53	28	-1	0	-38	-3	-4	-5	-39	-40	17	41

AREA B

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
1	-12	35	58	100	138	171	246	243	228	177	111	43	-22	-96	-146	-157	-166
2	-44	21	58	97	139	168	257	280	267	228	157	77	0	-81	-136	-157	-168
3	-73	2	56	106	170	219	308	371	349	305	230	128	31	-58	-119	-146	-149
4	-77	4	70	136	217	285	368	458	429	376	288	180	84	-4	-70	-111	-112
5	-98	-1	82	163	253	321	418	529	512	459	371	263	167	84	8	-46	-53
6	-121	-24	81	178	282	372	491	611	625	566	482	372	271	173	84	26	7
7	-122	-20	104	216	324	435	565	690	734	658	570	468	364	266	177	98	54
8	-109	-5	128	247	361	479	618	747	828	767	689	575	443	308	182	84	19
9	-67	13	141	263	383	515	685	837	1000	849	713	558	411	272	151	48	-20
10	-50	26	141	271	401	536	671	764	833	764	639	497	351	209	86	-17	-82
11	0	44	139	248	340	464	583	682	758	713	586	446	296	162	50	-36	-91
12	7	27	97	206	311	426	544	624	674	657	534	403	270	148	51	-25	-72
13	35	32	95	178	262	367	476	551	594	607	493	378	263	147	57	-14	-62
14	36	10	61	138	228	322	425	492	528	550	457	357	242	132	48	-16	-62
15	24	12	49	119	212	297	386	439	466	477	410	309	212	113	33	-34	-86
16	4	3	29	103	198	255	326	376	398	399	361	259	173	81	5	-61	-113
17	-4	12	9	75	163	216	275	320	344	331	307	207	123	40	-39	-96	-149

FIGURE 4-2. CORRELATION EXAMPLE, NO MODIFICATION ADDED

Area A and Area B have been chosen as representative of two possible extremes present in the data. Area A represents an area which appears to a viewer as perfectly flat, devoid of any activity. When a closer look is taken of the digital data, it becomes apparent that a noise background is present which is not apparent to the viewer. Because of this, the spectrum of frequencies present in the reference window is relatively flat. The correlation peak obtained is extremely narrow. This is due to the relative strength of the high frequency spectrum components.

Area B represents an area of the picture which appears to the viewer as very active. Such regions are found, for example, in the areas in which the cars are found. The correlation peak for this area is much broader than for Area A. This is due to the relative strength of lower frequency spectrum components present. In this case, the lower frequency components are much stronger than the higher frequency components.

4.1.2 Four Modification Examples

We would like to find modifications of the original aerial image which destroy the ability of conventional correlation techniques to determine match points. Toward this goal, four modifications have been tried.

4.1.2.1 Addition of Noise

For the first modification example, the original picture was modified with the addition of Gaussian noise. The ten reference windows were taken from the modified image and the search areas were taken from the original image. Figure 4-3 shows the results for Area A and Area B which are typical of the results for all ten cases. The standard deviation of the noise introduced was $\pm$ 40.

AREA A

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
1	75	[illegible]	-21	1	24	-7	-20	-29	-30	-56	-8	-23	-40	-12	-29.	27	20
2	-6[illegible]	-[illegible]	-30	57	10	0	24	-10	51	-9	-41	-64	-24	-7	-44	-20	13
3	[illegible]	[illegible]	13	-22	60	49	40	18	7	-7	6	64	8	2	-29	-21	-9
4	[illegible]	[illegible]	-12	16	-1	26	26	-5	11	53	57	-10	-37	-4	22	33	-5
5	-[illegible]	[illegible]	42	-12	-29	0	-1	-16	61	-19	91	45	38	77	41	95	26
6	[illegible]	[illegible]	-25	26	-18	33	71	7	-13	11	3	-26	35	33	20	30	31
7	[illegible]	[illegible]	74	23	54	21	38	62	55	47	35	81	82	17	31	30	41
8	21	14	7	43	27	53	105	94	88	71	35	-19	17	37	20	63	37
9	7[illegible]	69	48	17	15	4	117	210	[847]	214	89	-7	-8	14	29	79	26
10	[illegible]	66	3	55	47	0	46	103	78	97	83	68	17	51	50	33	-14
11	36	2	34	30	91	57	52	80	50	73	40	-16	50	38	63	90	-5
12	20	56	-3	4	23	-5	-13	10	-5	8	57	7	-43	17	-15	61	25
13	14	80	7	0	11	30	89	-9	38	-28	9	-29	-29	86	25	44	-12
14	-4	-8	23	-33	-85	-29	6	22	-9	0	-6	-35	-15	1	-32	12	0
15	15	-5	-19	-26	-24	-36	-15	-17	3	-34	69	52	34	-7	-1	-15	-16
16	5	12	-33	-25	-46	-50	34	-6	19	-1	-13	-31	-40	11	-12	5	-12
17	1	9	-26	-33	-41	-35	43	-24	14	-34	-7	16	-14	-53	-56	13	48

AREA B

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
1	-14	33	59	93	133	170	240	234	218	170	112	45	-19	-91	-133	-142	-155
2	-40	22	58	98	131	161	242	265	252	218	152	75	1	-80	-128	-147	-159
3	-73	4	54	99	160	210	297	359	339	296	226	124	31	-57	-114	-142	-147
4	-77	8	72	139	211	274	353	441	415	361	276	173	79	-10	-77	-108	-110
5	-89	1	85	161	242	311	406	510	494	444	359	257	160	78	7	-45	-52
6	-111	-22	83	181	278	362	480	593	610	549	465	359	261	166	75	23	7
7	-113	-12	103	217	320	431	553	671	715	639	548	455	354	258	169	93	48
8	-103	1	124	245	356	469	607	729	812	750	670	556	433	305	178	81	22
9	-63	20	140	259	380	510	675	823	[980]	832	699	547	405	271	156	50	-14
10	-39	29	140	270	394	530	660	752	815	748	626	488	346	211	93	-10	-75
11	2	47	142	248	336	453	570	673	747	706	578	441	296	165	57	-28	-81
12	24	34	104	213	311	422	537	621	669	653	528	399	270	151	58	-16	-61
13	41	42	102	187	265	367	469	545	589	600	489	374	266	150	61	-11	-57
14	41	20	72	147	226	320	416	482	516	539	449	349	238	130	46	-19	-63
15	36	25	53	131	221	294	381	429	458	465	402	299	205	109	32	-34	-84
16	25	17	35	106	201	255	318	363	388	387	354	249	168	77	0	-64	-110
17	8	32	14	75	164	211	270	306	333	319	296	196	115	40	-39	-96	-150

FIGURE 4-3. CORRELATION EXAMPLE, NOISE MODIFICATION ADDED

The standard deviations of the reference window data before the addition of noise were $\pm$ 63 and $\pm$ 198 for Area A and Area B, respectively.

Inspection of the two correlation peaks shown in Figure 4-3 shows that accurate matches can still be obtained. The correlation peaks maintain approximately the same width and are correctly centered. The main difference appears in the values obtained at the maximum. For Area A and Area B the value at the correlation peak maximum goes from 1.0 to 0.847 and 0.980, respectively. The value at the correlation peak maximum is dependent upon the signal to noise ratio. Area B represents a high signal to noise ratio case, and little change from a correlation maximum of one is observed. Area A represents a lower signal to noise ratio case, and a more significant change from one is observed in the value of the correlation peak maximum.

In general, the standard correlation approach to matching appears to handle the addition of noise very adequately.

4.1.2.2 Low Pass Filtering

Next, the original aerial image was modified by the use of a low pass filtering operation. A simple low pass filter was used which makes use of an averaging operation. Each pixel value was replaced by the average value from the five by five pixel area immediately surrounding it. For each of the ten correlation cases tried, the reference windows were taken from the modified image and the search areas were taken from the original image. Figure 4-4 shows the results from Area A and Area B which are typical of the results for all ten cases.

AREA A

0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
1	36	5	-14	-26	-41	-39	-28	-38	-48	-56	-61	-64	-67	-41	-41	-23	-6
2	41	17	-3	-14	-32	-26	-13	-8	-2	-15	-25	-26	-38	-42	-33	-4	20
3	38	26	7	4	-1	-7	2	8	24	17	3	2	7	12	23	56	81
4	4	10	0	9	3	1	2	20	38	28	24	25	30	41	61	86	97
5	37	34	31	32	29	25	31	46	64	64	73	79	90	101	116	120	121
6	32	47	46	52	50	56	60	78	102	96	97	100	122	138	164	151	131
7	47	62	52	60	77	109	245	287	332	313	286	156	138	143	163	141	125
8	59	60	63	65	93	135	282	332	372	348	312	173	132	125	129	104	82
9	95	101	109	116	130	166	318	360	391	370	327	174	129	121	122	109	98
10	101	105	105	102	113	151	298	335	365	336	277	124	69	67	77	87	77
11	111	113	106	96	104	135	281	310	332	289	236	89	46	47	48	64	60
12	113	114	104	92	92	106	120	130	139	114	89	72	49	47	40	43	31
13	96	74	57	52	57	68	60	60	81	55	46	44	26	18	7	12	15
14	87	62	53	14	11	18	39	43	57	34	25	24	1	-6	-19	-13	-16
15	65	38	2	-4	-2	2	28	34	44	15	4	-7	-19	-31	-31	-15	-2
16	46	20	-11	-21	-24	-16	6	15	28	9	-7	-24	-37	-44	-23	1	25
17	33	13	-12	-13	-43	-44	-19	-5	6	-1	-15	-28	-35	-25	1	18	36

AREA B

0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
1	-18	24	64	103	150	195	240	263	254	206	145	67	-15	-87	-138	-174	-195
2	-37	26	65	106	159	211	271	311	315	266	194	106	13	-66	-122	-161	-183
3	-61	12	70	127	193	260	337	389	401	350	266	166	66	-23	-86	-130	-152
4	-74	10	80	151	241	321	412	478	498	447	354	249	141	44	-25	-74	-98
5	-82	7	104	198	292	384	493	572	601	555	461	352	239	134	55	-2	-34
6	-90	7	117	229	338	447	574	668	707	669	578	463	343	226	130	58	15
7	-78	17	135	258	361	506	656	767	815	782	689	559	426	295	183	97	42
8	-56	32	155	284	415	552	716	831	881	850	755	615	470	326	200	102	35
9	-25	51	170	299	431	573	739	856	905	875	778	629	473	320	185	80	8
10	-5	63	172	293	429	573	732	845	889	858	758	607	445	289	151	45	-23
11	24	77	168	280	404	544	693	799	838	806	708	560	402	250	120	20	-44
12	47	78	147	245	362	493	624	721	759	732	642	510	362	220	99	6	-53
13	46	70	129	214	321	440	555	647	686	664	586	468	332	202	91	5	-50
14	48	63	110	181	278	385	487	568	609	595	529	428	307	187	85	3	-52
15	38	47	90	154	245	339	431	499	538	526	470	383	274	161	62	-18	-79
16	18	27	73	133	217	300	382	436	470	457	405	323	223	114	20	-55	-117
17	5	16	63	118	192	264	332	372	396	380	332	257	164	59	-27	-99	-160

FIGURE 4-4. CORRELATION EXAMPLE, LOW PASS FILTERING MODIFICATION ADDED

In all cases the match point found by the center of the correlation peak is still correctly positioned. However, in some cases the shape of the peak is substantially altered. The results for Area A are an example of this. With the addition of the low pass filtering, the peak becomes much broader and has a maximum at a value of only 0.391. Originally in this case, the correlation peak was very narrow because of the relative strength of the high frequency spectrum components. The low pass filtering removes the high frequency components from the reference window, and results in a much broader correlation peak.

For Area B, the original correlation peak is much broader corresponding with the dominance of low frequency components. The addition of the filtering operation has a much less dramatic effect on the correlation peak. It becomes slightly broader, and peaks at the value of 0.905.

Again, we find that the standard correlation approach to matching can handle modifications introduced through the use of a low pass filtering operation.

4.1.2.3 Nonlinear Intensity Transformation

For the next modification tried, a nonlinear transformation is applied to each intensity value. Figure 4-5 shows the nonlinear intensity transformation which has been used. With this transformation the output intensity values are exponentially related to the input values. For each of the ten correlation cases tried, the reference windows were taken from the modified image and the search areas were taken from the original image.

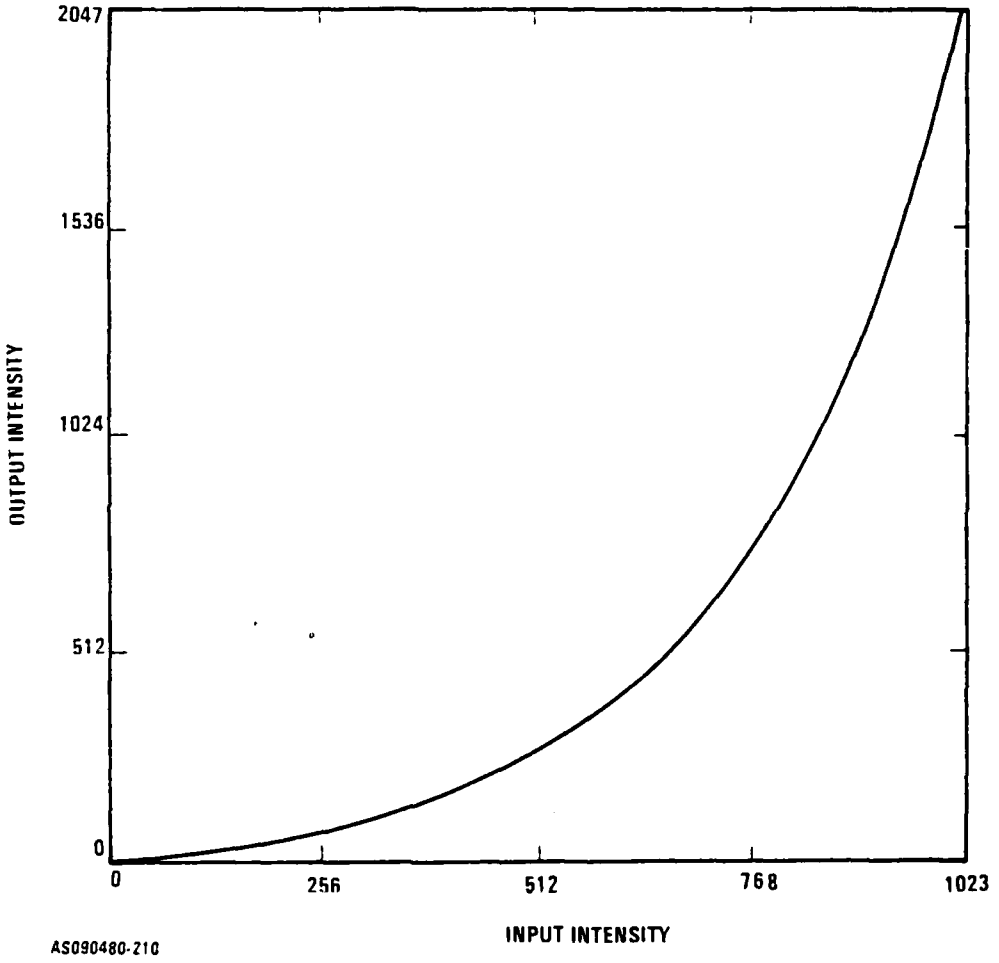


FIGURE 4-5. NONLINEAR INTENSITY TRANSFORMATION

Again, in all ten cases the match point found by the center of the correlation peak is still correctly positioned. In each case the correlation peak becomes slightly broader and the value at the maximum is slightly less than one. This can be seen in Figure 4-6 for Area 1 and Area 2. In these two cases the correlation peak occurs with values of 0.913 and 0.930, respectively.

Thus, the standard correlation approach to matching appears to handle cases of reasonable nonlinear modifications applied to the intensity levels. This particular property is due to the normalizing procedure used in the cross correlation product we have used. Of course, correlation will not work if the non-linear intensity transformation is dependent upon the scene content (the situation for IR imaging).

4.1.2.4 Edge Enhancement Operation

The final modification that has been used is an edge enhancement operation. Consider the three dimensional space made up of the two spatial dimensions corresponding with the pixel positions and the intensity values corresponding with the third dimension. In this space the sequence of intensity values from the original image define a surface. For each pixel position the gradient of the surface gives the magnitude and direction of the maximum slope associated with each pixel. The edge enhancement operator that has been used estimates the magnitude of gradient associated with each pixel position. The pixel values from a three by three neighborhood surrounding the pixel of interest are used in making the estimate. For the pixel position (i,j) let the neighborhood intensity values be denoted by:

AREA A

0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
1	79	-3	6	-33	-15	-13	-20	-25	-17	-39	8	-17	-18	-7	-10	27	-20
2	-44	-33	-25	58	14	-3	20	-2	-4	-55	-40	-46	-41	-30	-5	8	28
3	-13	26	12	-3	42	5	-9	-30	-9	23	34	43	58	48	-20	4	-11
4	74	25	-15	24	57	43	23	15	32	91	48	-44	-67	3	42	14	-7
5	-39	15	30	-11	-44	-7	30	14	69	0	81	66	50	33	28	83	29
6	-4	44	-54	-17	-75	-12	9	-16	-26	27	-3	-15	29	36	6	38	27
7	20	79	92	53	52	12	35	70	64	7	0	40	87	65	70	74	103
8	-12	11	20	6	34	50	43	89	73	56	44	11	19	43	51	94	64
9	57	34	6	17	-14	-9	103	266	413	245	86	2	-17	-15	46	72	45
10	46	69	21	49	29	3	43	96	62	102	75	52	33	32	51	40	-26
11	26	1	26	33	63	72	32	32	72	86	43	-4	20	16	59	55	27
12	47	47	-4	-10	9	-14	3	22	-17	21	61	32	-29	-6	-20	50	6
13	43	61	-10	12	26	65	115	-12	35	-10	-13	-21	10	46	22	39	-12
14	-5	2	27	-16	-59	-22	13	66	-8	-18	0	-19	-14	-9	-24	-23	-27
15	15	13	-17	-33	4	1	9	-14	-6	-7	91	63	63	-22	-25	-34	-40
16	47	29	5	2	-5	-22	-13	-11	31	-4	-22	-32	-22	25	-9	6	-34
17	-22	21	-24	-28	-38	-42	28	14	0	-23	8	6	16	-49	-59	-13	11

AREA B

0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
1	10	58	154	222	261	313	389	377	341	263	167	74	-8	-93	-156	-168	-163
2	-14	73	150	219	282	319	403	413	360	311	219	113	13	-86	-158	-179	-180
3	-39	46	135	214	294	356	442	489	452	381	281	152	32	-76	-155	-186	-183
4	-36	45	133	228	331	407	482	549	512	436	327	195	75	-38	-127	-179	-179
5	-46	39	141	239	343	418	497	594	568	495	378	241	121	15	-74	-139	-147
6	-63	20	134	236	348	436	536	639	634	559	457	324	193	75	-28	-97	-112
7	-64	22	140	256	375	480	589	690	711	621	510	387	264	154	51	-39	-82
8	-62	23	166	295	411	515	618	714	771	703	614	465	340	189	47	-61	-123
9	-7	62	174	256	414	531	674	800	930	785	634	458	295	144	16	-89	-159
10	-1	65	168	309	439	561	660	714	763	678	546	389	229	74	-56	-163	-226
11	23	90	169	275	366	476	563	625	677	617	482	329	171	25	-90	-182	-236
12	37	82	124	233	338	430	520	565	569	558	427	268	147	17	-80	-159	-202
13	55	74	125	194	280	370	457	496	509	503	386	271	148	29	-63	-137	-185
14	43	46	89	156	248	334	418	456	461	463	364	254	132	18	-70	-138	-182
15	24	45	82	142	245	311	395	426	425	414	329	213	107	10	-74	-143	-190
16	11	37	70	132	223	270	343	373	372	350	292	175	87	-5	-82	-149	-198
17	-5	34	49	114	145	241	300	336	338	305	267	156	63	-25	-104	-166	-215

FIGURE 4-6. CORRELATION EXAMPLE, NONLINEAR INTENSITY TRANSFORMATION MODIFICATION ADDED

$$\begin{matrix} x_{i-1,j-1} & x_{i-1,j} & x_{i-1,j+1} \\ x_{i,j-1} & x_{i,j} & x_{i,j+1} \\ x_{i+1,j-1} & x_{i+1,j} & x_{i+1,j+1} \end{matrix}$$

Then the edge enhancement operator that has been used is defined by:

$$s_x = (x_{i-1,j+1}+x_{i,j+1}+x_{i+1,j+1}-x_{i-1,j-1}-x_{i,j-1}-x_{i+1,j-1})/6 \quad (4\text{-}7)$$

$$s_y = (x_{i-1,j-1}+x_{i-1,j}+x_{i-1,j+1}-x_{i+1,j-1}-x_{i+1,j}-x_{i+1,j+1})/6 \quad (4\text{-}8)$$

$$s_{i,j} = \sqrt{s_x^2 + s_y^2} \quad (4\text{-}9)$$

$s_{i,j}$ gives an estimate of the magnitude of the gradient at the pixel position (i,j).

There are two ways to derive this estimate. If it is assumed that the nine points all are in a plane, then the least squares solution for the plane yields the above solution. Interestingly enough, the same result is obtained if the surface is assumed to be quadratic in its dependence on position. There are many edge enhancement operators that can be found in the digital image processing literature. The above estimate gives a reasonable compromise between accurately measuring sharp edges and providing some resistance to the effects of noise. Using fewer pixel values in the estimate more accurately describes sharp slope changes but is adversely affected by the presence of noise.

For each of the ten correlation cases tried, the reference windows were taken from the modified image and the search areas were taken from the original image. The results show that the application of the edge enhancement operator has completely destroyed the effectiveness of the correlation approach to matching.

Figure 4-7 shows the results for Area A and Area B. Figure 4-8 shows the distribution of the position of the correlation maximum for both the x and y directions for the ten correlation cases. These distributions are approximately uniform. As a check of this basic result, the larger reference window size was used. For a 64 by 64 pixel reference window the same basic result was obtained. Match points could not be determined.

This result is exactly what has been sought after. Figure 4-9 shows the edge enhanced image. To the human observer it is immediately obvious how to match the modified image to the unmodified one. For most of the class of dissimilar images which we would like to match this is the case. It is expected the differences present between these two images that prevent the correlation approach to matching from working are representative of the differences found in many of the dissimilar images that we would like to match.

4.2 The Texture Matching Approach

In this subsection, a simple texture matching approach of determining match points is demonstrated. The basic idea is to develop a descriptor that would be associated with each pixel that would be a measure of the surround of that pixel. This descriptor would describe the local scene content or texture. The procedure used is to replace each pixel from the images that are to be matched with this local measure of texture. In general, different replacement algorithms could be used for each image. Then conventional correlation techniques are used to determine match points. In the example presented in this subsection, the normalized cross correlation product is used for the correlation.

AREA A

0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
1	6	0	-32	-8	-5	0	27	15	-16	-70	-45	-41	0	24	-42	-12	-11
2	50	8	-15	-3	-22	-19	-38	-54	-72	-54	-25	-15	12	20	41	-6	30
3	22	-21	-3	-4	-6	0	-10	1	-17	-44	-35	-6	-10	-18	-13	37	59
4	17	-23	-17	-33	-46	-7	-9	-19	48	27	38	25	26	66	53	73	62
5	-31	-58	-56	-19	-49	-1	9	5	22	13	14	5	-6	16	5	11	34
6	-66	2	22	0	-33	11	25	69	30	17	0	0	13	-5	29	36	75
7	-36	-34	-4	-2	-39	-1	-32	12	10	22	-30	-34	10	49	61	15	5
8	5	37	79	47	6	25	64	124	93	144	20	16	25	34	33	39	54
9	-10	-4	-2	36	33	12	34	60	[43]	98	44	35	17	25	18	19	15
10	-3	6	13	27	48	54	69	103	125	(100)	75	45	19	35	39	43	18
11	21	23	51	0	-26	22	19	53	64	89	117	98	75	7	5	38	50
12	2	16	36	14	52	42	54	82	57	0	36	39	65	50	58	68	77
13	16	-6	11	-16	-5	26	18	57	43	23	43	65	60	-6	48	-1	20
14	11	-11	47	18	73	74	14	41	3	22	20	16	33	18	45	37	35
15	8	1	7	-19	-33	-14	-16	12	-40	32	-26	24	21	38	19	20	50
16	22	-24	-42	-45	-37	20	-33	-26	-59	39	18	67	31	26	2	-12	35
17	-21	45	-8	-22	-22	-34	-63	-43	-46	-14	-37	-20	-11	-35	26	3	40

AREA B

0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
1	373	335	322	345	365	369	382	339	223	159	79	26	-21	-57	-121	-180	-200
2	298	344	344	351	350	368	383	353	249	158	97	41	2	-27	-92	-167	-196
3	318	341	365	372	375	386	391	369	280	167	106	51	16	-9	-71	-158	-192
4	279	347	338	390	387	388	365	343	262	136	87	52	27	11	-33	-114	-152
5	272	351	378	406	392	384	344	328	264	152	114	97	87	78	43	-34	-94
6	284	374	421	423	413	417	393	378	321	219	176	165	144	122	84	9	-57
7	294	411	444	(445)	434	446	447	408	341	230	161	151	127	108	59	-12	-88
8	296	407	429	436	425	424	417	344	268	176	90	71	48	36	7	-46	-120
9	252	381	427	445	430	419	404	318	[220]	130	40	30	18	10	-4	-41	-112
10	228	356	395	442	410	336	353	254	157	75	-8	-22	-14	-10	-24	-66	-145
11	222	367	440	433	376	351	324	243	131	59	-4	-31	-31	-45	-71	-119	-183
12	215	323	407	413	400	355	313	240	113	34	-38	-59	-103	-111	-127	-157	-202
13	194	283	410	430	419	350	285	210	74	-15	-84	-143	-136	-139	-155	-184	-221
14	179	267	397	418	359	329	256	167	59	-35	-99	-166	-170	-167	-185	-210	-242
15	174	245	366	411	395	317	229	151	40	-54	-114	-175	-193	-199	-218	-229	-247
16	165	210	329	393	372	292	201	113	13	-84	-130	-181	-220	-216	-228	-236	-261
17	147	169	280	364	348	269	178	93	9	-96	-142	-169	-233	-225	-239	-251	-269

FIGURE 4-7. CORRELATION EXAMPLE, EDGE ENHANCEMENT MODIFICATION ADDED

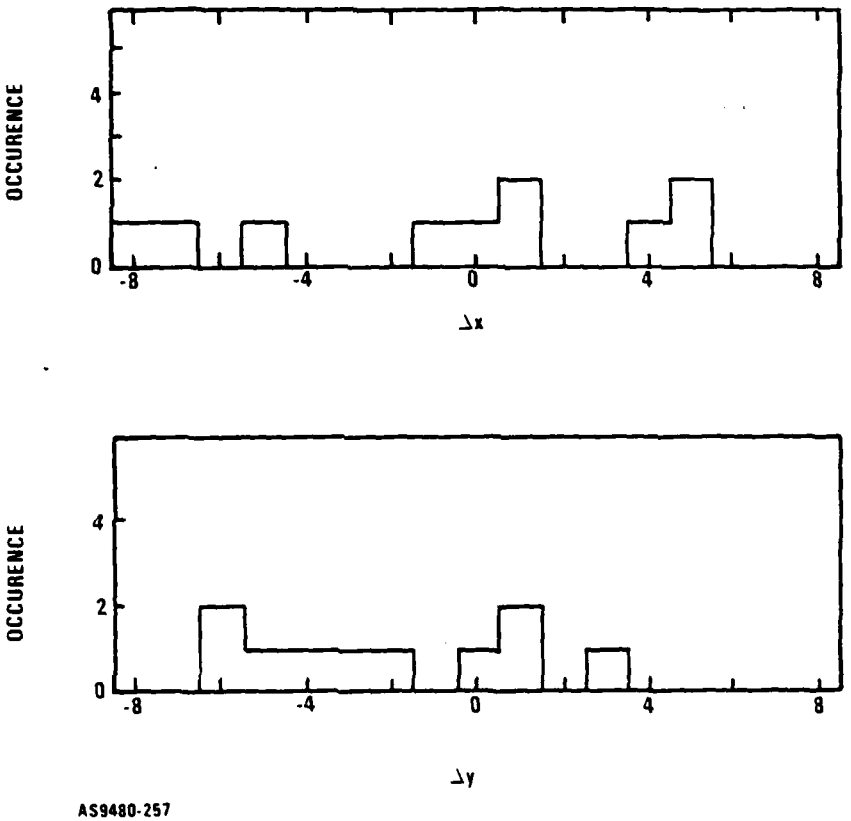


FIGURE 4-8. ERROR IN CORRELATION PEAK POSITION, EDGE ENHANCEMENT MODIFICATION EXAMPLE

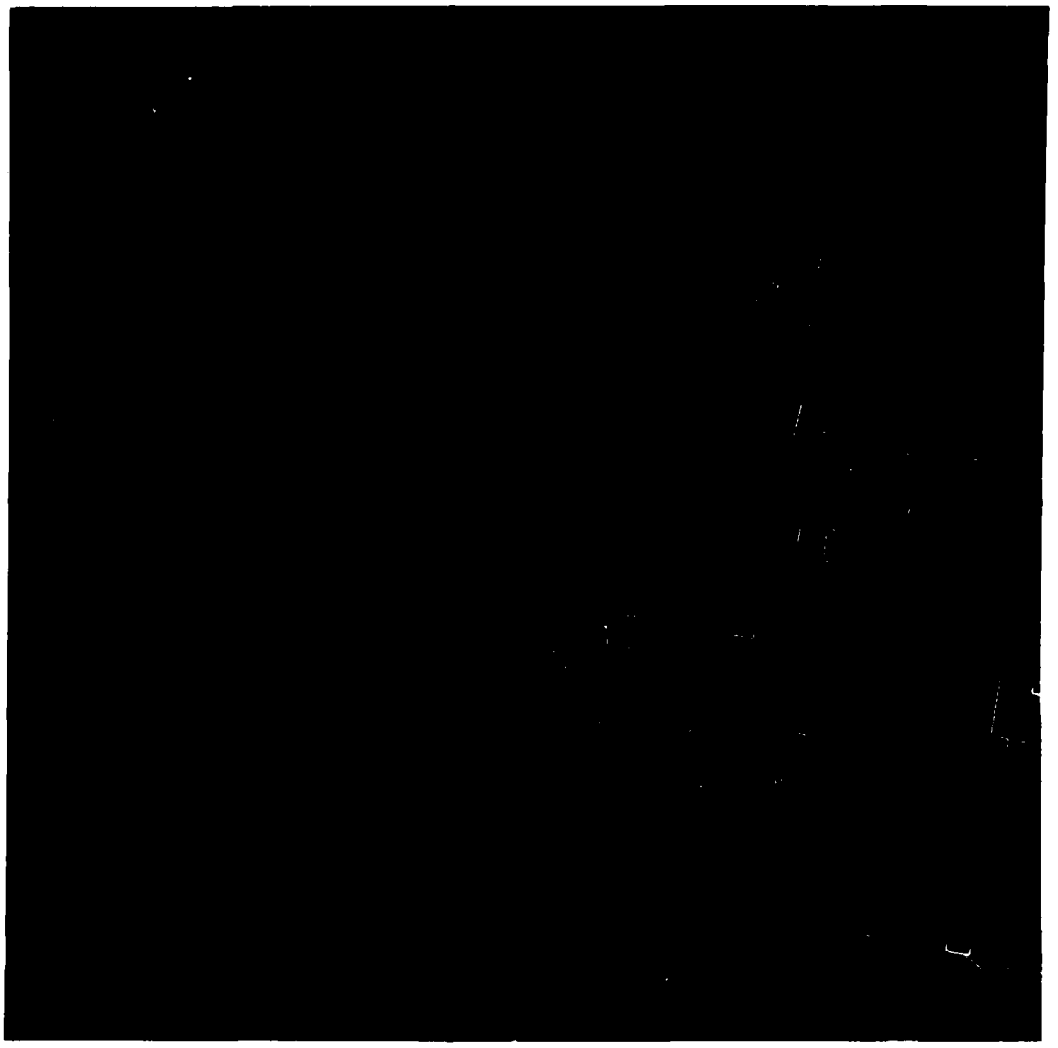


FIGURE 4—9.

**SAMPLE AERIAL IMAGE
MODIFIED BY AN EDGE ENHANCEMENT OPERATION**

In order to demonstrate the approach, the results from the previous subsection are used. Here, it was demonstrated that after the application of an edge enhancement operation, normal correlation techniques no longer are successful in determining match points. This is true despite the fact that a human observer would have no trouble in matching the two images. Thus, we will attempt to match the original aerial image with the image modified by an edge enhancement operation using the texture matching approach.

4.2.1 Texture Matching Example

A very simple quantity is used here as a measure of local texture. For each pixel, the pixel values in a four by four neighborhood surrounding the pixel of interest are used. These pixel values are averaged to find a mean intensity, and then the root-sum-square deviation about the mean is calculated. Using the same notation as was used to define the edge enhancement operator, this transformation is defined as:

$$\bar{x} = \frac{1}{16} \sum_{k=-1}^{+2} \sum_{l=-1}^{+2} x_{i+k,\, j+l} \tag{4-10}$$

$$x_{i,j} = \sqrt{\sum_{k=-1}^{+2} \sum_{l=-1}^{+2} (x_{i+k,\, j+l} - \bar{x})^2} \tag{4-11}$$

This measure of local activity or texture is referred to as the local busyness measure. In regions near edges or in regions of rough texture its magnitude is large. In calm or flat areas its magnitude is small.

The procedure used is to transform both the original aerial image and the image distorted by the edge enhancement operation. Although the same transformation is used here for each image of the pair, in general, the transformations could be different. All pixels are transformed into the edge busyness. Then the normalized cross correlation product is used to determine match points. The reference windows are selected from the modified image and the search areas are selected from the original image. Again a reference window of 32 by 32 pixels is used and a search area of 48 by 48 pixels is used. For the ten matching cases that are tried, exactly the same areas are selected that were used in the correlation experiments in the previous subsection.

Figure 4-10 shows the correlation results for Area A and Area B. As can be seen, a well defined correlation peak appears in both cases. The center of the peak is within one pixel of the exactly correct match position. Of the ten examples tried this is true in all but one of the cases. In more than half the cases, the value at the correlation peak is 0.8 or higher. In three cases the value at the correlation peak is approximately 0.55. Thus, the transformation of both the images into the simple texture measure has permitted the determination of match points in most cases to within one pixel accuracy.

Figure 4-11A shows the texture matching results for the one case in which an accurate match was not obtained. In this case the correlation peak consists of a long diagonal ridge with values at the top of the ridge very nearly the same. In this direction the position of the match is poorly determined resulting in the bad match obtained. To obtain a match on this region of the image it was found necessary to enlarge the reference window size. Figure 4-11B shows the results when a 64 by 64 pixel reference area is used. A correct match is obtained.

AREA A

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
1	-91	-91	-113	-92	-94	-117	-114	-117	-103	-80	-61	-28	-10	10	23	30	52
2	-...	-71	-39	-31	-36	-47	-53	-56	-47	-31	-20	5	29	48	45	27	24
3	0	3	6	25	37	23	14	8	11	19	31	66	85	104	86	46	25
4	15	35	56	89	87	93	72	65	63	65	72	96	110	116	94	54	31
5	4	32	83	133	154	156	139	121	117	112	106	105	108	121	100	69	44
6	-25	15	81	149	193	229	266	307	332	329	278	223	165	120	100	83	60
7	-55	-13	61	149	236	314	396	463	529	518	420	303	190	104	74	67	50
8	-129	-72	13	124	208	403	533	656	713	661	536	369	212	94	48	40	21
9	-169	-13	-37	67	254	425	576	714	769	716	557	369	191	69	4	-8	-19
10	-24...	-170	-78	4...	2...	401	546	677	714	661	497	305	130	0	-61	-76	-74
11	-141	-163	-79	20	101	215	340	521	551	473	335	182	53	-75	-131	-142	-122
12	-14...	-133	-55	29	142	212	279	324	311	201	100	62	-36	-104	-142	-141	-108
13	-...	-7...	-...	5	6...	87	100	9...	6...	37	-6	-44	-77	-94	-99	-82	-41
14		-1...	-...	3	3	33	29	12	-17	-35	-50	-50	-45	-27	-6	20	53
15	9...	4	1...	2	3	-5	-24	-47	-75	-81	-73	-46	-16	20	59	91	115
16	12...	1...	1...	-1...	-3...	-4...	-67	-84	-108	-107	-82	-34	6	50	102	144	171
17	1...	71	1...	-3...	-74	-9...	-11...	-131	-145	-129	-85	-26	11	47	86	124	163

AREA B

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
1	3...	3...	3...	455	47...	5...	5...	5...	5...	5...	444	41...	42...	43...	40...	33...	242
2	41...	41...	4...	47...	5...	5...	5...	5...	5...	5...	48...	45...	45...	47...	45...	39...	303
3	4...	4...	4...	4...	5...	5...	6...	6...	6...	6...	5...	4...	49...	50...	49...	43...	34...
4	4...	4...	4...	4...	5...	5...	5...	6...	6...	6...	5...	5...	5...	5...	5...	45...	37...
5	5...	5...	5...	5...	5...	5...	6...	7...	7...	6...	6...	5...	5...	5...	5...	48...	41...
6	4...	4...	5...	5...	5...	5...	6...	7...	7...	7...	6...	5...	5...	5...	5...	52...	45...
7	4...	4...	5...	5...	5...	5...	7...	7...	8...	7...	7...	5...	5...	5...	5...	52...	45...
8	4...	4...	5...	5...	5...	6...	7...	8...	8...	8...	7...	6...	5...	5...	5...	52...	45...
9	4...	4...	4...	4...	5...	5...	6...	7...	8...	8...	7...	6...	5...	5...	5...	49...	42...
10	3...	4...	4...	4...	4...	5...	6...	7...	8...	8...	7...	6...	5...	5...	5...	4...	35...
11	3...	4...	4...	4...	4...	4...	6...	7...	8...	8...	7...	5...	4...	4...	3...	3...	31...
12	3...	4...	4...	4...	3...	4...	5...	7...	8...	7...	6...	5...	3...	3...	3...	3...	27...
13	3...	3...	4...	4...	3...	3...	5...	6...	7...	7...	6...	5...	3...	2...	2...	2...	23...
14	2...	3...	4...	4...	3...	3...	4...	6...	7...	7...	6...	5...	4...	2...	2...	2...	21...
15	2...	3...	4...	4...	3...	3...	4...	6...	7...	7...	6...	5...	4...	3...	2...	2...	20...
16	2...	3...	4...	4...	3...	3...	4...	5...	6...	7...	6...	5...	4...	3...	2...	2...	20...
17	2...	3...	4...	4...	3...	3...	4...	5...	6...	6...	6...	4...	4...	3...	2...	2...	21...

FIGURE 4-10. TEXTURE MATCHING EXAMPLE, ONE-BY-ONE PIXEL REPLACEMENT RESOLUTION

A) 32 BY 32 PIXEL REFERENCE WINDOW

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
1	620	[illegible]	723	707	619	481	336	237	160	107	-19	-163	-278	-331	-326	-295	-260
2	569	[illegible]	732	756	719	611	459	314	225	171	83	-52	-197	-302	-341	-327	-296
3	502	617	708	764	775	717	592	431	295	221	162	61	-81	-223	-316	-340	-323
4	434	551	663	746	791	785	707	566	405	263	214	148	33	-114	-249	-327	-342
5	380	478	594	704	778	813	790	692	541	384	272	206	131	6	-144	-269	-335
6	351	413	520	638	738	804	827	788	677	521	368	263	199	114	-21	-169	-288
7	333	365	464	565	672	766	824	838	785	662	499	349	255	191	93	-48	-198
8	299	336	431	466	585	698	787	840	841	774	637	471	332	249	181	73	-76
9	227	291	383	386	485	607	719	801	843	833	749	604	443	317	244	172	51
10	125	216	285	332	399	508	628	735	809	844	820	723	574	418	306	238	158
11	21	114	213	282	336	415	528	649	752	821	846	809	701	549	401	299	231
12	-62	-20	124	219	287	345	432	551	673	772	834	847	798	679	526	386	292
13	-111	-55	31	137	229	293	354	450	575	697	790	842	847	783	658	506	371
14	-132	-103	-42	52	156	243	301	388	473	602	719	804	845	842	768	638	483
15	-110	-102	-34	-27	72	176	255	311	385	499	628	739	818	855	839	754	614
16	-106	-124	-120	-85	-10	91	189	261	318	402	523	651	756	830	859	832	733
17	-67	-112	-121	-113	-71	6	107	200	267	329	422	547	674	776	839	860	817

B) 64 BY 64 PIXEL REFERENCE WINDOW

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
1	30[illegible]	351	384	376	356	320	287	265	243	198	133	62	0	-44	-60	-67	-76
2	388	350	393	413	414	389	350	307	270	227	171	101	29	-31	-60	-73	-83
3	257	336	399	438	461	456	424	368	306	247	193	132	61	-9	-52	-74	-84
4	230	315	394	451	492	510	494	437	354	266	197	142	83	16	-34	-64	-77
5	207	293	364	432	491	532	543	505	422	315	219	158	115	61	6	-33	-59
6	182	260	311	385	462	535	584	586	521	422	302	215	167	122	62	9	-35
7	152	201	259	334	428	516	628	676	654	566	433	311	234	182	118	53	-9
8	114	172	227	297	385	521	657	743	756	690	557	412	301	229	162	92	21
9	69	144	204	266	350	488	632	746	816	751	639	497	370	276	196	126	52
10	15	101	178	237	298	411	561	693	756	749	666	542	415	310	214	140	78
11	-34	47	135	203	250	326	453	591	663	681	632	541	437	341	242	163	109
12	-74	-11	69	144	172	239	328	441	559	589	579	526	461	391	303	218	158
13	-109	-56	-4	61	109	173	232	322	428	508	541	530	495	456	387	297	217
14	-121	-136	-69	0	71	135	186	254	347	439	501	523	519	502	450	364	271
15	-123	-119	-90	-47	22	103	169	214	278	356	428	472	493	500	476	411	323
16	-111	-123	-110	-36	-33	45	120	172	217	272	340	396	439	466	469	434	365
17	-135	-116	-121	-111	-76	-12	64	127	170	205	254	312	365	409	432	427	387

FIGURE 4-11. TEXTURE MATCHING EXAMPLE, FAILURE CASE

4.2.2 Resolution Factors

Several areas have been investigated to learn about some of the factors affecting the resolution obtained in using the texture matching approach. The most interesting results were obtained when different pixel replacement resolutions were studied. In the previous subsection each and every pixel from the 1000 by 1000 pixel images were replaced by their local busyness measure as defined by Equation 4-11. This replacement resolution is referred to as one-by-one replacement resolution. Alternately, it is possible to replace only every other pixel in every other line. This gives a 500 by 500 local busyness representation of the original images. This is referred to as two-by-two replacement resolution. Replacing every fourth pixel in every fourth line gives four-by-four replacement resolution.

It is important to distinguish what is being done here from a similar but different process. One could take the original image and abstract every other pixel from every other line. Then the local busyness could be determined using this subset of the original pixels. This is not what is being done here. Rather, local busyness is just being calculated for every other pixel in every other line exactly as orginally defined.

The use of two-by-two and four-by-four replacement resolution have been studied. Again ten matches are attempted for both resolutions. In each case the same areas of the images are used for the reference windows and for the search areas. For two-by-two replacement resolution, the original 32 by 32 pixel reference window becomes 16 by 16 elements wide, the original 48 by 48 pixel search area becomes 24 by 24 elements wide, and the

size of the correlation space becomes 9 by 9 instead of 17 by 17 For the four-by-four replacement resolution, the original 32 by 32 pixel reference windows becomes 8 by 8 elements wide, the original 48 by 48 pixel search area becomes 12 by 12 elements wide, and the size of the correlation space becomes 5 by 5.

Figure 4-12 shows the matching results for two-by-two replacement resolution for Area A and Area B. For a correct match the correlation peak should occur at the fifth column of the fifth row. Each row or column off this corresponds with a two pixel displacement. Figure 4-13 shows the matching results for four-by-four replacement resolution for Area A and Area B. For a correct match the correlation peak should occur at the third column of the third row. Each row or column off this corresponds with a four pixel displacement.

We would like to compare the resolution in match point determination for the three replacement resolution approaches (one-by-one, two-by-two, and four-by-four). In order to obtain the position of the correlation peak to the accuracy of a fraction of a pixel, a simple fitting procedure has been used. The x and y directions are treated independently. In each direction a three point fit to a parabola is performed to determine the peak position.

Table 4-I summarizes the results of this procedure. In most cases, the accuracy of the match is correct to within a fraction of a pixel in both directions. For one case (case 5), a reasonable match was not obtained and this case has not been used. For the nine cases left the average accuracy in the x and y directions has been obtained from the square root of the sum of the errors squared. The average radial error in the match point determination

AREA A

	1	2	3	4	5	6	7	8	9
1	-125	-100	-55	-96	-104	-62	10	73	89
2	-1	56	91	53	25	57	119	105	45
3	16	142	208	236	274	235	170	111	54
4	-60	122	285	467	613	452	178	64	39
5	-202	-30	261	578	(755)	531	150	1	-18
6	-239	-119	126	390	510	323	20	-113	-79
7	-110	-102	17	90	65	0	-87	-146	-71
8	49	-37	-36	-50	-131	-123	-33	30	91
9	135	-18	-96	-136	-199	-147	-14	91	167

AREA B

	1	2	3	4	5	6	7	8	9
1	345	441	457	538	535	439	411	413	249
2	368	479	491	608	625	529	473	492	328
3	405	516	515	681	732	614	493	520	396
4	441	528	501	702	816	702	530	552	443
5	442	507	479	718	(889)	744	471	461	378
6	395	467	415	642	843	699	393	352	305
7	349	437	376	541	753	652	372	269	224
8	325	426	373	472	695	653	411	256	195
9	273	409	373	421	632	640	451	270	195

Figure 4-12. TEXTURE MATCHING EXAMPLE, TWO-BY-TWO PIXEL REPLACEMENT RESOLUTION

AREA A

	1	2	3	4	5
1	135	-77	-87	-3	62
2	-33	269	446	104	26
3	-203	313	[715]	68	-91
4	128	-86	125	-108	-159
5	178	-122	-222	-57	21

AREA B

	1	2	3	4	5
1	419	329	544	380	266
2	441	393	767	414	429
3	441	412	[859]	354	351
4	414	349	765	258	221
5	317	267	633	312	185

FIGURE 4-13. TEXTURE MATCHING EXAMPLE, FOUR-BY-FOUR PIXEL REPLACEMENT RESOLUTION

TABLE 4-I.

SUMMARY, MATCH POINT ERRORS, TEXTURE MATCHING EXAMPLES

Case \ PRR a)	Δx			Δy		
	4X4	2X2	1X1	4X4	2X2	1X1
1(A)	-0.58	-0.12	0.02	0.74	0.27	0.00
2(B)	-0.12	0.08	0.37	0.02	-0.23	-0.50
3	-0.14	0.46	0.88	0.40	0.44	0.26
4	-0.18	-0.87	-0.69	-0.60	-1.11	-0.78
5	---	---	---	---	---	---
6	-0.16	-0.35	-0.69	-0.39	-0.72	-0.84
7	-0.67	-0.41	-0.79	-1.16	-0.52	-0.89
8	0.46	0.07	0.50	-0.47	-0.49	-0.50
9	0.30	-0.01	-0.22	0.24	-0.44	-0.81
10	-0.08	-0.34	-1.07	-2.44	-2.44	-1.44
RSS	0.36	0.39	0.76	0.99	0.99	0.84

a) PRR = Pixel Replacement Resolution

is then obtained by adding in quadrature the average error in the x and y directions. The results yield $\pm$ 1.04, $\pm$ 1.04, and $\pm$ 1.08 for the 4 x 4, 2 x 2, and 1 x 1 replacement resolutions, respectively. In other words, all three replacement resolutions determine the match point to approximately the same accuracy.

This is an important result. The use of the four-by-four replacement resolution requires one-sixteenth fewer calculations than the one-by-one replacement resolution case. This result is not completely unexpected. The calculation of local busyness from one pixel to the next is highly redundant. However, it was expected that there would be some fall off in match point resolution, in going from two-by-two to four-by-four replacement resolution.

One other factor affecting match point resolution was investigated. The effect of increasing the reference window size from 32 by 32 to 64 by 64 pixels was studied. It was found that the match point resolution was improved by approximately a factor of two.

4.3 Some General Conclusions

A few general conclusions can be drawn from these simple experiments conducted to demonstrate the texture matching approach to matching dissimilar images. In most parts of most images, the power present in the lower frequency components far exceeds that found in the higher frequency components. Using the conventional approach to correlation matching, this causes the low frequency components to dominate the selection of the match point. In dissimilar images of the type we would like to match, it is usually

found that the lower frequency components from one image to the next do not agree. Thus, to have any hope of success in matching dissimilar images, the technique must not use information related to the low frequency behavior of the images. Note that in using the texture matching approach that has been suggested, this will always be the case. If, in general, each pixel is replaced by some reasonable measure of local texture, then the low frequency components of the image are effectively discarded.

Next consider the size of the neighborhood used to determine the local texture measure. In the previous subsection, a four-by-four pixel neighborhood was used to determine the local busyness texture measure. What would be the effect on match point resolution if the size of this neighborhood is made larger? In the last subsection we demonstrated that it is not necessary to replace every pixel by the texture measure. Further, not replacing every pixel does not effect the match point resolution obtained. This is due to the high correlation present in the calculations of the texture measures from one pixel to the next. As the size of the neighborhood used is increased this correlation should become even stronger. Thus, it is suggested that as size of the neighborhood used is increased, the match point resolution obtained will correspondingly increase (get worse). For this reason, it is suggested that the size of the neighborhood used in determining the texture measure be limited to four-by-four, eight-by-eight, or, perhaps, sixteen-by-sixteen pixels about the pixel of interest. In some cases larger areas may be needed, but, in general, it is expected that the optimum area size will be quite small. It is further

suggested that as a general rule that the pixel replacement resolution to be used correspond with half the distance used in defining the neighborhood. In other words, for a four-by-four neighborhood, a two-by-two pixel replacement resolution be used or for a eight-by-eight neighborhood, a four-by-four pixel replacement resolution be used. Closer spaced pixel replacement resolution would only increase the computational load with no increase in match point resolution. Broader spaced pixel replacement resolution must eventually degrade the match point resolution. Time permitted for this program did not permit a detailed study to confirm these suggestions.

5.0 TEXTURE MATCHING REFINEMENTS

In this section refinements to the basic texture matching approach to image matching are discussed. Algorithms are suggested that could be coded in FORTRAN IV and tested on the DIAL facility at USAETL. Before discussing these refinements, the basic approach is summarized.

The two images to be matched are first corrected for all known geometric distortions using the initial knowledge of the taking conditions (see Section 2). As part of this process, resampling of the images is performed if necessary. This resampling process aligns the coordinate axes used in the two images and provides for equal sample spacing for the two images. It is assumed that the dominant effect of inaccuracies in the taking conditions give a translational mismatch between the two images. Each of the images is independently transformed on a pixel by pixel basis. Each pixel intensity value is replaced by a local measure of the local scene content (local texture). A reference window is chosen in one image and a larger search area is chosen in the other image. The match point is determined using the normalized cross correlation product (see Section 4.1). The match point is determined by the position of the correlation peak. A threshold can be set using the value of the correlation parameter at the peak to determine if a good match has been found. This process can be repeated across the two images to obtain match points as frequently as desired.

In the first subsection a few simple transformations are discussed for image matching. In the next subsection texture measures which can be derived from local spectral analysis are discussed. The local spectral analysis is performed with the use

of the two dimensional discrete Fourier transform. This is a powerful approach and generates a broad class of texture measures. It becomes clear in this subsection that to describe texture in detail requires more than just a single texture measure. In the next subsection the texture matching technique is formally defined, and this technique is generalized to handle multiple measures of texture. The next subsection discusses a variety of special problems which can interfere with texture matching. Methods are discussed for dealing with data noise characteristics, different detector spatial acceptances, drastically different intensity distributions, and coherency effects found in dealing with imaging radars.

5.1 Some Simple Transformations

The two transformations discussed in this subsection are not strictly speaking transformations which measure local texture. However, they are transformations which allow image matching between a broader class of images than would be true for conventional correlation matching. The two transformations replace the original image by either a high pass frequency filtered version of the image or an edge enhanced version of the image.

5.1.1 High Pass Filtering

In Section 4.3 it was pointed out that in most of the dissimilar images of the type we would like to match, the lower frequency components do not agree. Therefore, one approach to the problem is to remove these components through the use of two dimensional high pass filtering. An example of where this approach might be helpful would be two overhead images taken from the same perspective but at different times of the day. An appropriate high-pass filter can be developed with standard signal analysis tools.

5.1.2 Edge Enhancement Operation

A slightly broader class of images could be matched if images are first processed by an edge enhancement operator such as was defined in Section 4.1.2.4. This operator measures the magnitude of the gradient of the intensity level surface. Notice that for an operator of this type, an edge that goes from a low intensity level to a high intensity level will yield the same response as an edge that goes from a high intensity level to a low intensity level. This is because the magnitude of the gradient is used. In this way it differs from the use of a high pass filter and permits a slightly broader class of images to be matched. For actual implementation it is suggested that the specific edge enhancement operator defined in Section 4.1.2.4 be used. As was pointed out, this operator provides a reasonable compromise between accurately measuring sharp edges and providing some resistance to the effects of noise.

5.2 Texture Measures from Spectral Analysis

In this subsection texture measures which can be derived from local spectral analysis are considered. For each pixel, a neighborhood is defined about which you wish to measure parameters associated with the local texture. For the neighborhood, the local spectral components are determined using the two dimensional discrete Fourier transform. For computational simplicity, the neighborhood size dimensions are usually chosen to be some power of two pixels wide. For example, the neighborhood size might be four-by-four, eight-by-eight, or sixteen-by-sixteen pixels wide. Once the frequency components are determined, the texture measure is determined from these component values. The general approach here uses the square root of the power present in similar groups of frequency components. For example, the power present in the low frequency

terms or the high frequency terms could be used. It quickly becomes evident that to describe texture in detail requires more than one texture measure. In order to accommodate more than one texture parameter, the general matching procedure that has been presented needs to be generalized. This is discussed in Section 4.3.2.

To be able to group together similar spectral components in a reasonable manner requires some knowledge of the properties of the discrete Fourier transform. These properties are discussed here first using the one dimension discrete Fourier transform.

Given a sequence of N equally spaced intensity values denoted by $x_0, x_1, \ldots, x_{N-1}$, its discrete Fourier transform is given by

$$X_u = \frac{1}{\sqrt{N}} \sum_{j=0}^{N-1} x_j \, e^{\frac{-2\pi i \, ju}{N}} \tag{5-1}$$

where $i = \sqrt{-1}$. The inverse transform is given by

$$x_j = \frac{1}{\sqrt{N}} \sum_{u=o}^{N-1} X_u \, e^{\frac{2\pi i \, ju}{N}} \tag{5-2}$$

It should be noted that there is not a universally accepted form of these equations. Some times a reversal of signs is used in the kernels. Some authors prefer to place all scaling constants in the inverse transform equation. The normalization used here leads to a particularly simple form for Parseval's relationship.

$$\sum_{i=0}^{N-1} x_i^2 = \sum_{u=0}^{N-1} |X_u|^2 \tag{5-3}$$

Note that the spectral components as defined are complex. At first glance it appears that from N independent spatial components we have produced 2N independent frequency components. Actually this is not true. Some of the components are only real. Others are closely related. The relationship between spectrum components is most easily expressed if the periodic nature of the spectrum components is understood. Normally, the frequency components for $u = 0, 1, \ldots, N-1$ are calculated. This is not the only choice. Equation 5-1 has the property that the output is periodic in N, i.e.

$$X_u = X_{u+N} \tag{5-4}$$

for any u. It is convenient for our purposes to calculate the frequency components for $u = -N/2+1, \ldots, -1, 0, 1, \ldots, N/2$. In terms of these components the relationships that reduce the number of independent frequency components to N are:

$$\mathrm{Im}(X_0) = \mathrm{Im}(X_{N/2}) = 0 \tag{5-5}$$

$$X_u = X_{-u}^{*},\ u = 1, 2, \ldots, N/2-1 \quad (* = \text{conjugate}) \tag{5-6}$$

This particular choice of the range of u used in finding the frequency components is very important to the discussion to follow. With this choice, the frequency of the wave associated with the frequency component X_u is proportional to the absolute value of u. The spectral components X_2 and X_{-2} are associated with a wave of twice the frequency of components X_1 and X_{-1}, etc. We will be using the two dimensional generalization of the discrete Fourier transform to produce frequency components $X_{u,v}$. With this particular choice of the range on u and v, it is relatively easy to identify closely related frequency terms. With the more conventional choice this task becomes very awkward.

The use of the discrete Fourier transform can be extended in a straigth forward manner to two dimensions. In practice the two dimensional form is calculated by a sequence of applications of the one dimensional transform. The input data is arranged in a square matrix. Each column is transformed using the one dimensional transform. The rows of the resulting matrix are then each transformed using the one dimensional transform. The output are the two dimensional frequency components $X_{u,v}$. Alternately, the rows can be transformed first, followed by the columns.

The transform can be implemented using the well known fast algorithm, the fast Fourier transform (FFT). Using the FFT, the number of additions and multiplications become proportional to NlogN instead of N^2. For two dimensions, the number of additions and multiplications is proportional to $2N^2\log N$ instead of N^4.

The general prescription used to develop texture measures, groups together like frequency components. These are combined to give a texture measure using the general prescription

$$t = \sqrt{\sum_{u,v} \left| X_{u,v} \right|^2} \qquad (5\text{-}7)$$

where the sum over u,v includes all the like frequency components that have been selected. Note that with this prescription all phase information is discarded.

It is interesting to point out that the local busyness texture measure is of this form. This is the texture measure that was used in Section 4.2 to demonstrate the texture matching approach. This measure is in fact equal to the square-root of AC

energy contained within the block. In other words, to construct this texture measure (t_B), the above sum is taken over all u,v except (u,v) = (0,0). The equivalence of this definition of local busyness to the definition used in Section 4.2.1 can be demonstrated using Parseval's relationship (Equation 5-3).

It should be noted that the term $X_{0,0}$ is never used in defining any of the texture measures. This is the term which is proportional to the average intensity value of the block.

Texture measures defined in this manner can be combined to form new measures. For example, suppose the texture measures t_X and t_Y have been defined. Here t_X is a measure of strength of frequency components which are only a function of the x direction. t_Y is a measure of the strength of frequency components which are only a function of the y direction. Then a new texture measure $t_{X/Y}$ could be defined by dividing t_X by t_Y. This would give you a relative measure of the strength of waves which are a function of the x direction only to waves which are a function of the y direction only. Alternately, t_X and t_Y could be normalized by the local busyness measure, i.e.

$$t'_X = t_X/t_B, \quad t'_Y = t_Y/t_B \tag{5-8}$$

In the X direction, for example, this would give a ratio of the strength of waves which are a function of the x direction only to the overall strength of all waves present.

First, specific texture measures are suggested for neighborhoods defined to be four-by-four or eight-by-eight pixels wide. Then, a refinement is suggested for calculating the frequency components. The procedure suggested helps to reduce unwanted effects generated by the very strong low frequency components present in most images.

5.2.1 Four-by-Four Pixel Neighborhood

For a four-by-four neighborhood, the sixteen complex frequency coefficients are:

$$\begin{matrix} X_{2,-1} & X_{2,0} & X_{2,1} & X_{2,2} \\ X_{1,-1} & X_{1,0} & X_{1,1} & X_{1,2} \\ X_{0,-1} & X_{0,0} & X_{0,1} & X_{0,2} \\ X_{-1,-1} & X_{-1,0} & X_{-1,1} & X_{-1,2} \end{matrix}$$

In this form it is easy to group together similar frequency terms and form texture measures.

For a first example, the low or high frequency terms can be grouped together to form a measure of the strength of the low or high frequency terms. These are defined by

$$t_L = \sqrt{|X_{-1,-1}|^2 + |X_{0,-1}|^2 + |X_{1,-1}|^2 + |X_{1,0}|^2 + |X_{1,1}|^2 + |X_{0,1}|^2 + |X_{-1,1}|^2 + |X_{-1,0}|^2} \quad (5\text{-}9)$$

$$t_H = \sqrt{|X_{2,-1}|^2 + |X_{2,0}|^2 + |X_{2,1}|^2 + |X_{2,2}|^2 + |X_{1,2}|^2 + |X_{0,2}|^2 + |X_{-1,2}|^2} \quad (5\text{-}10)$$

The procedure used is a simple case of a general procedure of grouping together frequency components using rings defined about (0,0) in frequency space. This is further illustrated in the next subsection.

Alternatively, the strength of waves which are only a function of the x direction, the strength of waves which are only a function of the y direction, and the strength of waves which are a function of both directions can be measured. These are defined by:

$$t_X = \sqrt{|X_{0,-1}|^2 + |X_{0,1}|^2 + |X_{0,2}|^2} \quad (5\text{-}11)$$

$$t_Y = \sqrt{|X_{-1,0}|^2 + |X_{1,0}|^2 + |X_{1,1}|^2} \quad (5\text{-}12)$$

$$t_{XY} = \sqrt{|X_{-1,-1}|^2 + |X_{1,-1}|^2 + |X_{2,-1}|^2 + |X_{2,1}|^2 + |X_{2,2}|^2 + |X_{1,1}|^2 + |X_{1,2}|^2 + |X_{-1,1}|^2 + |X_{-1,2}|^2} \quad (5\text{-}13)$$

5.2.2 Eight-by-Eight Pixel Neighborhood

For an eight-by-eight neighborhood, there are 64 complex frequency coefficients. These are symbolically represented in Figure 5-1. In this figure the low, medium and high frequency terms have been grouped together using rings superimposed on the frequency components. The rings are defined by $\sqrt{u^2 + v^2} = 0.5$, 1.5, and 2.5. The components used to define the strength of the low, medium and high frequency components are listed in Table 5-I. These three texture measures are denoted by t_L, t_M, and t_H.

Figure 5-2 shows alternative groupings of similar frequency terms. Here pie shaped regions about the (0,0) element are used to define three texture measures. These are referred to as t_X, t_Y, and t_{XY}. The components used in each of these texture measures are listed in Table 5-I.

These six measures should be more than sufficient to describe the texture of an eight-by-eight neighborhood. Endless alternative descriptions could be formed by using ratios of the measures or by normalizing by the total AC energy (t_B). In any case, it should not be necessary to use more than a half dozen or so measures.

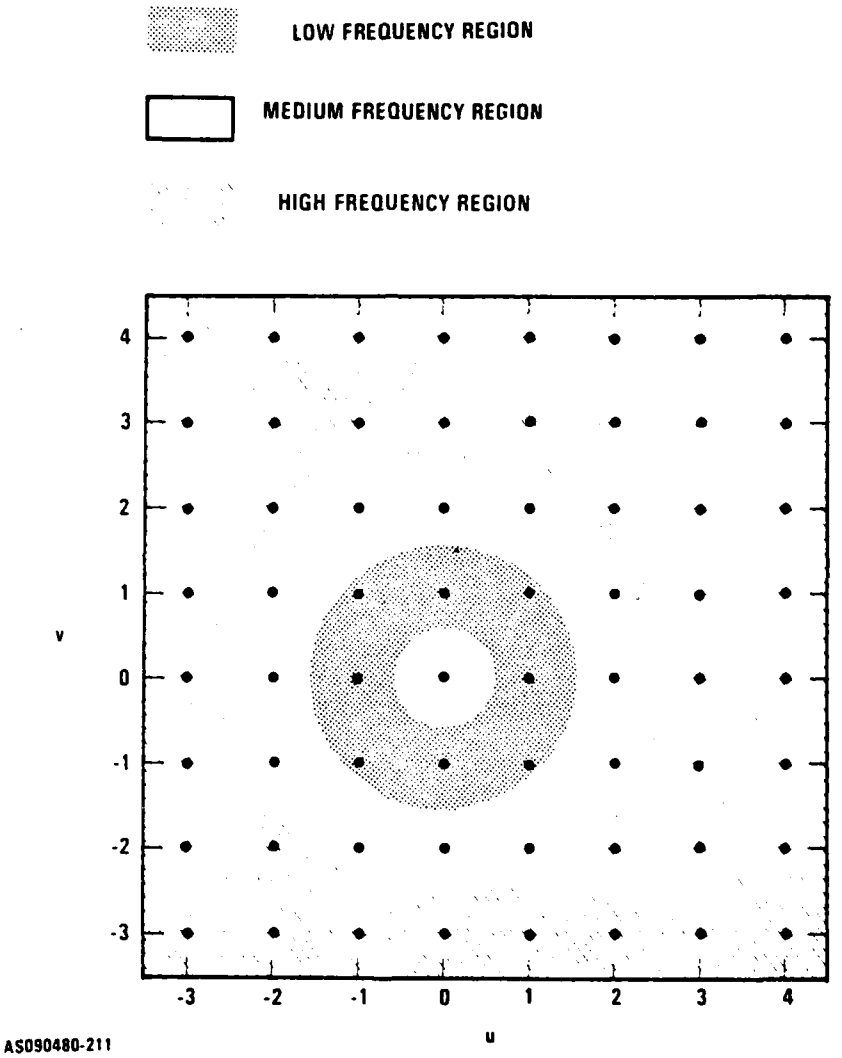


FIGURE 5-1. DIVISION OF FREQUENCY COMPONENTS DEFINING THE LOW, MEDIUM, AND HIGH FREQUENCY TEXTURE MEASURES

TABLE 5-I.

TEXTURE MEASURE COMPONENTS, 8X8 NEIGHBORHOOD CASE

Texture Measure	Frequency Components
t_L	(1,-1),(1,0),(1,1),(0,1),(-1,1),(-1,0),(-1,-1),(0,-1)
t_M	(2,-1),(2,0),(2,1),(1,2),(0,2),(-1,2),(-2,1),(-2,0),(-2,-1), (-1,-2),(0,-2),(1,-2)
t_H	(4,-3),(4,-2),(4,-1),(4,0),(4,1),(4,2),(4,3),(4,4),(3,-3), (3,-2),(3,-1),(3,0),(3,1),(3,2),(3,3),(3,4),(2,-3),(2,-2), (2,2),(2,3),(2,4),(1,-3),(1,3),(1,4),(0,-3),(0,3),(0,4), (-1,-3),(-1,3),(-1,4),(-2,-3),(-2,-2),(-2,2),(-2,3),(-2,4), (-3,-3),(-3,-2),(-3,-1),(-3,0),(-3,1),(-3,2),(-3,3),(-3,4)
t_X	(4,-1),(4,0),(4,1),(3,-1),(3,0),(3,1),(2,0),(1,0),(-1,0), (-2,0),(-3,-1),(-3,0),(-3,1)
t_Y	(1,-3),(1,3),(1,4),(0,-3),(0,-2),(0,-1),(0,1),(0,2),(0,3), (0,4),(-1,-3),(-1,3),(-1,4)
t_{XY}	(4,-3),(4,-2),(4,2),(4,3),(4,4),(3,-3),(3,-2),(3,2),(3,3) (3,4),(2,-3),(2,-2),(2,-1),(2,1),(2,2),(2,3),(2,4),(1,-2), (1,-1),(1,1),(1,2),(-1,-2),(-1,-1),(-1,1),(-1,2),(-2,-3), (-2,-2),(-2,-1),(-2,1),(-2,2),(-2,3),(-2,4),(-3,-3),(-3,-2), (-3,2),(-3,3),(-3,4)

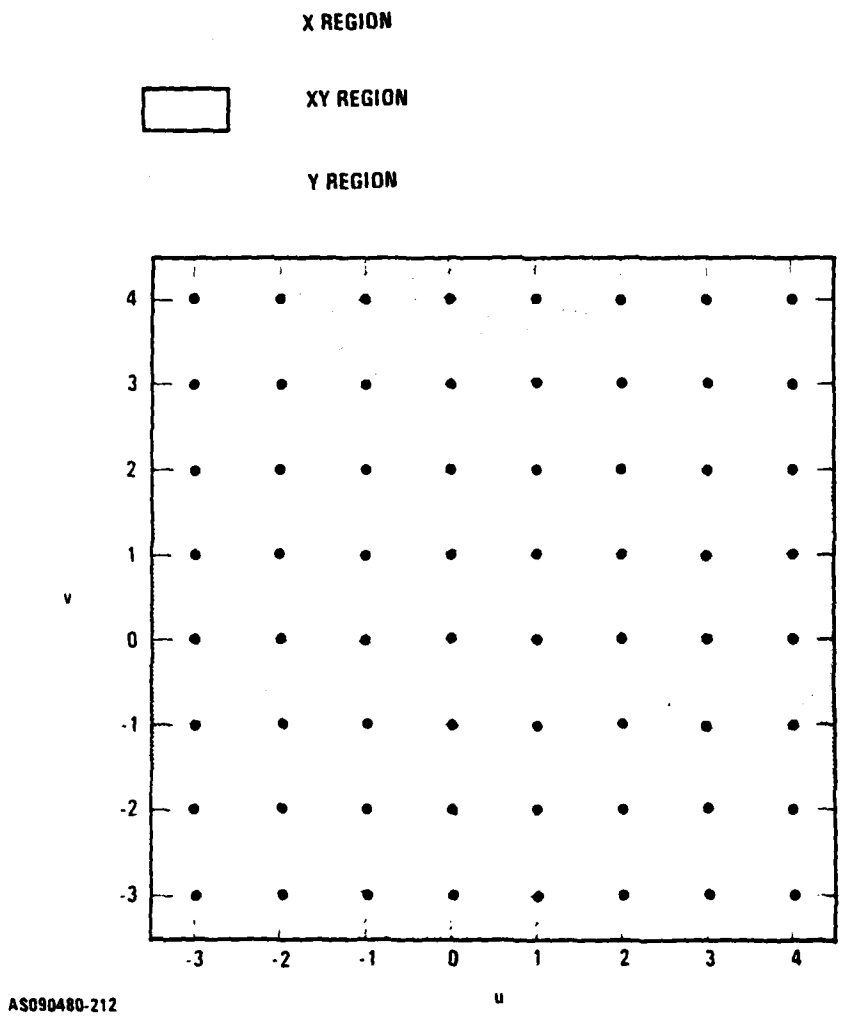


FIGURE 5-2. DIVISION OF FREQUENCY COMPONENTS DEFINING THE X, XY, AND Y TEXTURE MEASURES

If it was desired to describe texture more finely, the frequency components could be simultaneously divided by the radial and pie shaped regions shown to give nine texture components.

Note that to define texture measures by grouping together similar frequency terms, it has not been necessary to know the N relationships which reduce the 64 complex terms to 64 independent numbers. These relationships are useful, however, for the implementation of the calculation of the frequency components. For convenience in expressing these relationships, N/2 is denoted by M. These relationships are:

$$Im(X_{0,0}) = Im(X_{M,0}) = Im(X_{0,M}) = I_m(X_{M,M}) = 0 \qquad (5\text{-}14)$$

$$X_{u,v} = X^*_{-u,-v} \qquad u,v = 0,1, \ldots M\text{-}1 \qquad (5\text{-}15)$$

$$X_{u,M} = X^*_{-u,M} \text{ and } X_{M,u} = X^*_{M,-u} \qquad u = 1,2, \ldots,M\text{-}1 \qquad (5\text{-}16)$$

The use of these relationships reduces the number of one dimensional discrete Fourier transforms that must be performed from 2N to approximately 3/2 N. Figure 5-3 helps illustrate this for the eight-by-eight neighborhood case. This figure shows an example of a minimum set of components that must be explicitly calculated. The remaining components can be found from the above relationships. Eight one dimensional transforms would be required on the eight columns, but only five one dimensional transforms would be required on the rows instead of eight.

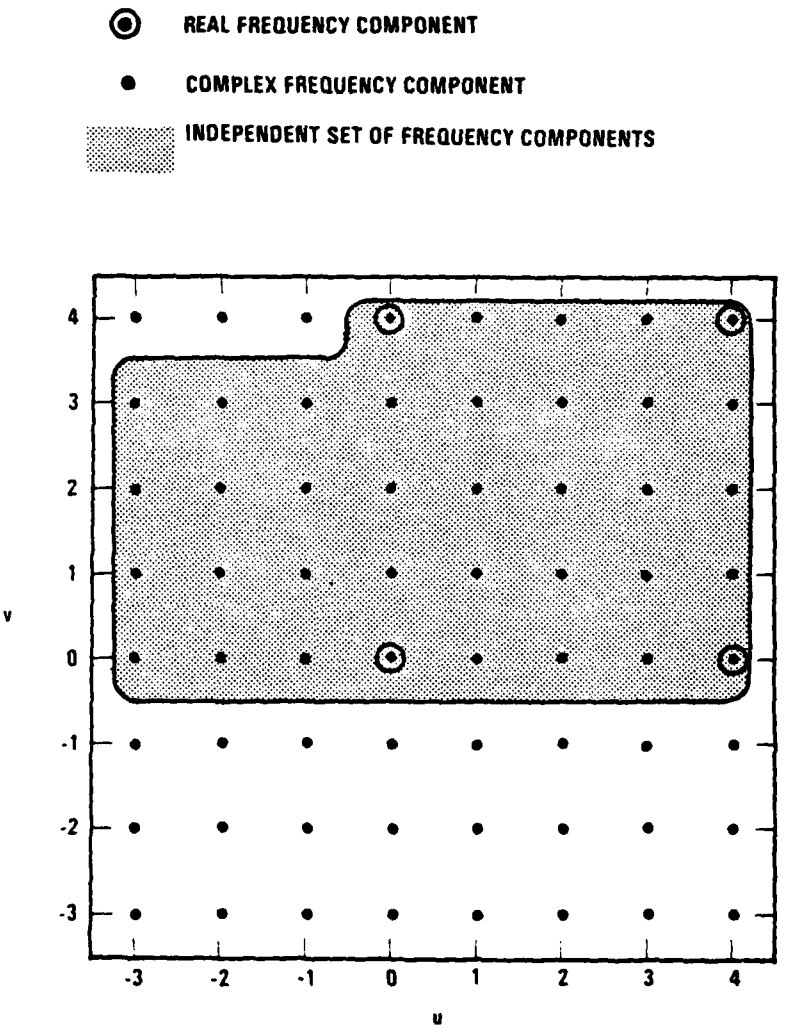


FIGURE 5-3. DEFINITION OF AN INDEPENDENT SET OF FREQUENCY COMPONENTS

5.2.3 Low Frequency Background Subtraction

In most images the very lowest frequency components have by far the most power. Between dissimilar images of the type to be matched, the low frequency behavior from one image to the next is usually very different. Because of their relative strength and because of the properties of the discrete Fourier transform, these very low frequency components can strongly interfere with the texture measures which have been defined in this subsection. This problem is illustrated in Figure 5-4. Here we consider what happens when a ramp signal is texture analyzed. From a texture point of view the signal is essentially flat, devoid of any texture. Thus, the texture measures should give null readings.

Unfortunately, this is not what happens. Assuming a block size of four pixels, the signal is divided into blocks and each block is analyzed separately. The Fourier transform of the block gives the spectrum of the discrete periodic signal generated by infinitely replicating the four pixel values in the block. The periodic nature of the discrete sequence represented by the discrete Fourier transform can be shown from Equation (5-2). The discrete periodic sequence resulting is shown in Figure 5-4. Calculation of the spectral components gives non-zero high frequency terms, and as a result the texture of the block is misrepresented.

The solution of this problem is straight forward. The data should be high pass filtered before the data is divided up into blocks and analyzed. The frequency cutoff should be at 1/N times the sampling frequency.

As a method of implementing this filter, the following method is suggested. First, the block means are determined.

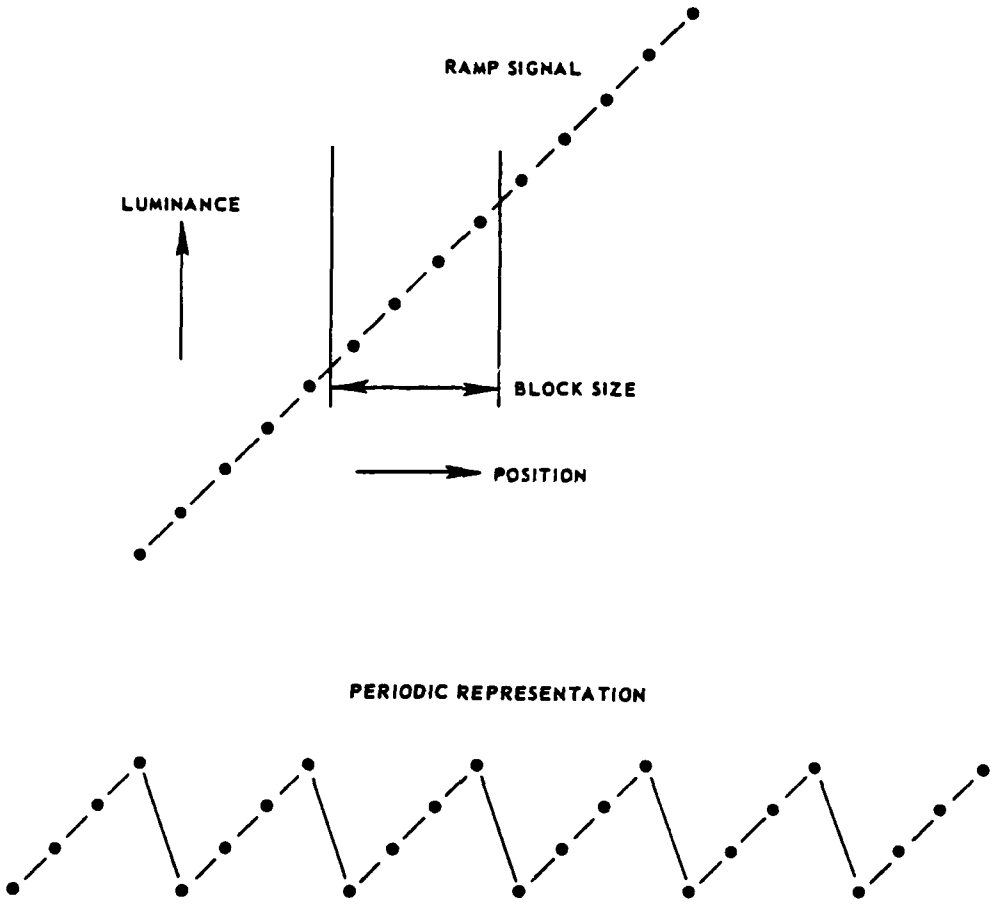


FIGURE 5-4. LOW FREQUENCY BACKGROUND SUBTRACTION EXAMPLE

Then values between the blocks means at points corresponding to the pixel positions are determined using interpolation. Simple linear interpolation should be sufficient to drastically reduce the problem. This interpolated signal is used as an estimate of the unwanted low frequency components and is subtracted from the original signal. Analysis of the residual signal is then performed as originally described.

Consider what happens when this approach is applied to the ramp signal example. Subtraction of estimated low frequency background results in a null signal out. The texture measurements will then properly reflect the flat texture present.

5.3 Formal Definition, Texture Matching

The texture matching procedure which has been described is formally summarized in this subsection. This is first done for the case of a single texture measure. Next, this procedure is generalized to handle multiple texture measures. Again for convenience one dimensional expressions are used.

5.3.1 Single Texture Measure

A reference window is selected from one image to be matched, and a search area is selected from the other image. Let the pixel replacement resolution be given by R. Then in pixel intensity space these areas are represented by:

reference window: x_i, $i = 1,2,\ldots,R*N$

search area: y_i, $i = 1,2,\ldots,R*M$, $M>N$.

From the pixel intensity measurements, the texture measure is determined for each Rth pixel. Let the texture measures derived for the reference window and the search area be denoted by:

reference window: s_i, $i = 1,2,\ldots,N$

search area: t_i, $i = 1,2,\ldots,M$, $M>N$.

Let A be the pixel neighborhood size used to determine the texture measure. Then each texture value is determined using A pixel values. Specifically

$$s_i = f(x_{(i-1)*R+1}, x_{(i-1)*R+2}, \ldots, x_{(i-1)*R+A}) \tag{5-17}$$

$$t_i = f(y_{(i-1)*R+1}, y_{(i-1)*R+2}, \ldots, y_{(i-1)*R+A}) \tag{5-18}$$

Then the normalized cross correlation product is formed between the texture measures for the reference window and the texture measures for the search area.

$$\rho_j = \frac{\frac{1}{N}\sum_{i=1}^{N} (\hat{s}_i \cdot \hat{t}_{i+j})}{\sigma^s \cdot \sigma_j^t} \qquad j = 1,2,\ldots,M-N+1 \tag{5-19}$$

with $$\hat{s}_i = s_i - \bar{s}, \qquad \bar{s} = \frac{1}{N}\sum_{i=1}^{N} s_i, \tag{5-20}$$

$$\hat{t}_{i+j} = t_{i+j} - \bar{t}_j \qquad \bar{t}_j = \frac{1}{N}\sum_{i=1}^{N} t_{i+j}, \tag{5-21}$$

and $$\sigma^s = \sqrt{\frac{1}{N}\sum_{i=1}^{N} (s_i - \bar{s})^2} \tag{5-22}$$

$$\sigma_j^t = \sqrt{\frac{1}{N}\sum_{i=1}^{N} (t_{i+j} - \bar{t}_j)^2} \tag{5-23}$$

The match point is selected from position of the correlation peak. For good matches, the correlation peak value will approach one. The parameters that must be experimentally determined for the matching of any two types of dissimilar images include:

a) the single type of texture measure to be used,

b) the size of the reference window and search area to be used,

c) the size of the neighborhood used to determine the texture measure,

d) and the pixel replacement resolution to be used.

If a single texture measure is to be used, the local busyness measure as defined by Equation 4-11 should be an excellent choice.

5.3.2 Multiple Texture Measures

To apply multiple texture measures, independent matching is performed for each of the measures used. Then an overall correlation parameter is used. Let B give the number of correlation measures used. Then the overall correlation parameter is defined by:

$$\rho_i = \frac{1}{B} \sum_{u=1}^{B} \rho_i^u \qquad (5\text{-}24)$$

The superscript u refers to each of the multiple texture measures used. The match point is determined by the peak of this distribution. The normalization by B allows the identification of good matches as peak values which approach one.

If it is possible to determine that some texture measures are more important than others, then a weighting procedure is easily defined. For example the overall correlation parameter could be defined by:

$$\rho_i = \sum_{u=1}^{B} \omega^u \rho_i^u, \tag{5-25}$$

with

$$\sum_{u=1}^{B} \omega^u = 1. \tag{5-26}$$

Here ω^u is a weighting factor of less than one associated with each of the texture measures. The normalization of the weighting factors allows the identification of good matches as correlation peak values which approach one. Methods for determining appropriate weighting factors are beyond the scope of this report.

5.4 Miscellaneous Corrections

In this subsection a variety of special problems which can interfere with texture matching are discussed. Methods are discussed for dealing with data noise characteristics, different detector spatial acceptances, drastically different intensity distributions, and coherency effects found with imaging radars. Detailed algorithms for dealing with these problems are not, in general, presented. The emphasis here is to point out the potential existance of these problems and to suggest lines along which the problems could be dealt with.

5.4.1 Noise Characteristics

Image data is frequently corrupted by significant amount of noise. This noise can interfere with the image matching process. We will consider here only the most commonly found type of noise, that of white, Gaussian noise. Consider an example in which the two images to be matched have differing levels of Gaussian noise. Assume that each intensity measurement from one image has

a noise standard deviation component which is twice as large as that in the second image. Consider what happens in a region of the scene which is of constant intensity if not corrupted by noise. Using the texture matching approach, the texture measures derived from spectral analysis from this region for the first image will be on the average twice as large as texture measures from this region in the second image. For a match attempt with a reference window which includes this region, the match will be adversely affected because of this. A straight forward solution to this problem exists. Noise should be added to the second image to bring it up to the level of the first image. This paradoxial solution gives texture measures in the region in question which are now on the average of equal magnitude.

5.4.2 Spatial Acceptance

Differences in spatial acceptances of the dissimilar images to be matched will interfere with the determination of corresponding points using texture matching. Differences in spatial acceptance causes each of the images to be blurred by different amounts. This causes the texture at corresponding points to be different from one image to the next. Again the solution to this problem is fairly straight forward.

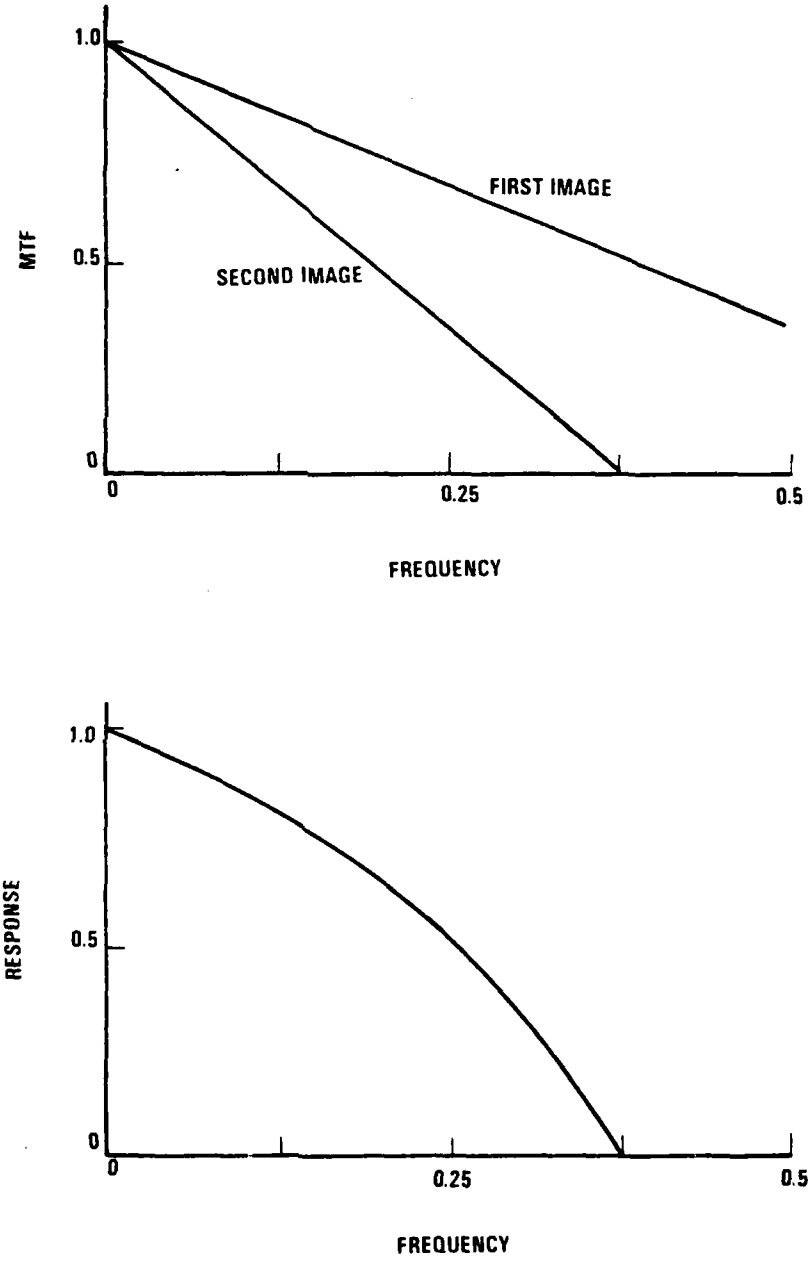
For simplicity consider the following example. A single aerial camera has been used to produce two images of the same area. The only difference between the two images is the altitude used. The second image is taken from an altitude which is twice as high as the altitude used for the first image. After sampling both images with what correspondes to equal sample spacing on the ground, the two images will not be quite identical. The second image will be blurred with respect to the first. In practice it

should still be possible to obtain match points with methods that have been described. However, it is possible to improve the matching process.

Improved results are obtained by adding additional blurring to the first image. The blurring can be accomplished with the use of an appropriate low pass filter. The required properties of the low pass filter, can be easily determined if the spatial acceptances for the two images are known. This is illustrated for the example being considered in Figure 5-5. The spatial acceptances for the two images are shown in the form of their modulation transfer functions. A modulation transfer function gives the attenuation of spatial waves as a function of frequency. Frequencies have been expressed by normalizing by the sampling frequency. The frequency response of the required filter is given by the ratio of the acceptance of second image to the acceptance of the first image as a function of frequency. The required response for the example is shown in Figure 5-5.

5.4.3 Intensity Distributions

In general, the matching procedure that has been described should be quite insensitive to broad differences in the magnitude and the shape of the intensity distributions of the two images to be matched However, with the matching of an optical image with a radar image extreme differences can be found in their intensity distributions. In general, the intensity distribution of an optical image is relatively flat. On the other hand, the intensity distribution of a typical radar image consists of an extremely large clutter peak at very low intensities plus tail made up of the intensities of a



AS090480-214

FIGURE 5-5. EXAMPLE MATCHING SPATIAL ACCEPTANCE WITH LOW PASS FILTERING.

few very bright targets. Well over half of the intensity values will frequently fall in the clutter peak. Typical distributions are illustrated in Figure 5-6.

Under these circumstances it is suggested that improved matching results could be obtained by modifying the radar image intensity distribution with the use of an intensity transformation. The transform could be of the form

$$I' = aI^b \tag{5-27}$$

where b is less than one. Figure 5-6 shows a modified intensity distribution of the original radar distribution that could be obtained by a transformation of this form.

5.4.4 Coherency Effects

The matching of high resolution radar images gives a severe test to any image matching technique. The coherent nature of imaging radar produces targets with a high frequency scintillation component. These waves shift in position with the slightest change in taking parameters. Because of this, conventional correlation matching procedures will frequently not work on radar images taken with nominally the same taking parameters. For this particular case, the texture matching procedure should overcome this difficulty. However, between imaging radars of different resolutions or between a radar image and an optical image, coherent effects present a difficult problem.

One suggestion to deal with this problem is to treat the high frequency coherent components as noise. Given that the high frequency components of one of the images are dominated by noise, matching using texture matching will be improved by low pass filtering both the images. The low pass filter should cutoff at the frequency at which noise starts to dominate.

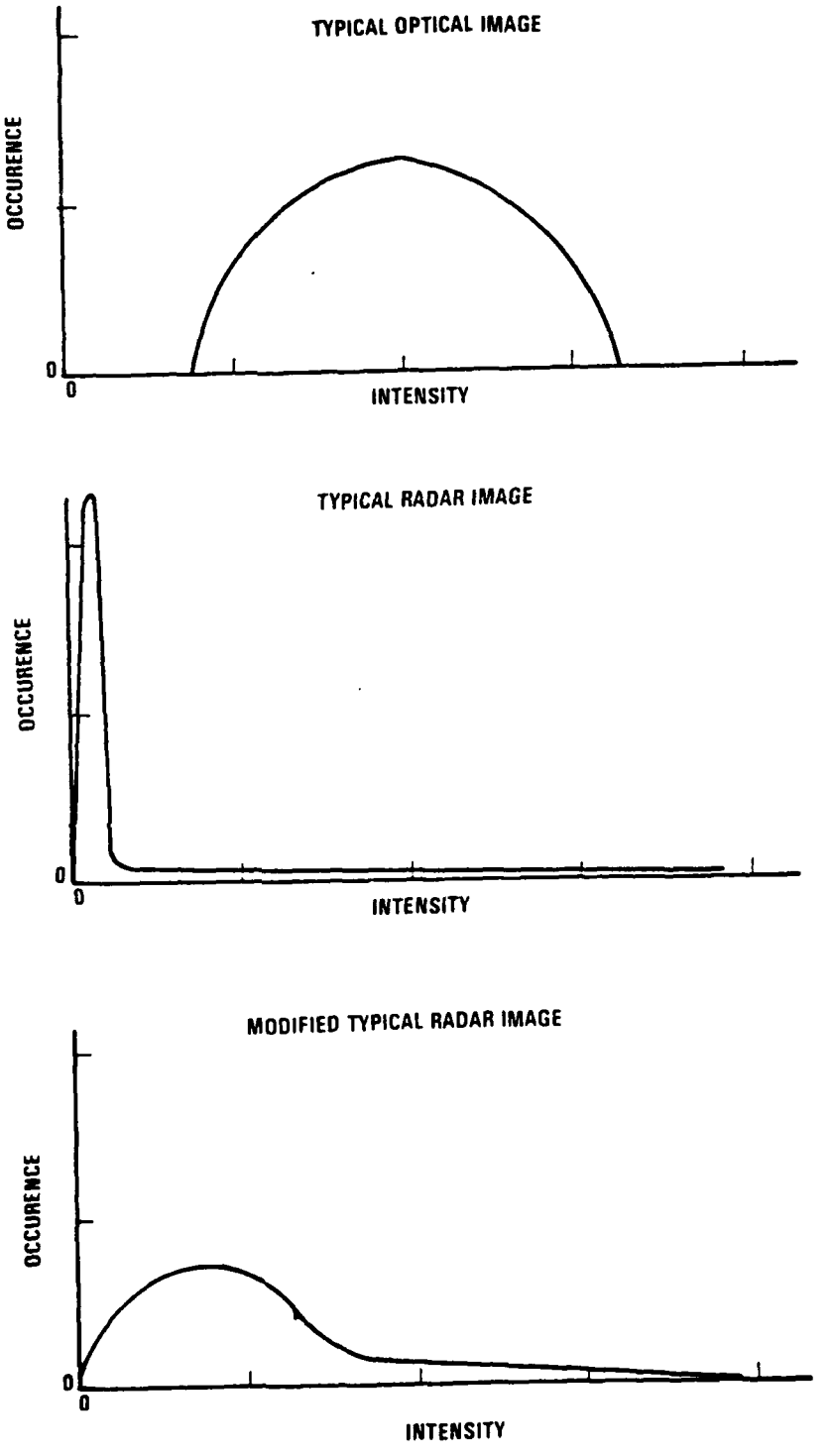


FIGURE 5-6. TYPICAL INTENSITY DISTRIBUTIONS

5.5 Match Point Accuracy

In general, match point accuracy will vary as match points are obtained across a pair of images. In some areas of a scene it may not be possible to obtain well defined match points, while in other areas highly accurate match points will be obtained. There are several approaches which can be taken to predict or evaluate the accuracy of a possible match point. In general, the more active a local area of the scene is, the more accurate the match point determinations will be. In an area completely devoid of any activity, match point determination will not be possible. In an area dominated by sharp edges or other sharply defined structures, highly accurate match points will be obtained. This fact can be used to select areas of the scene which give the most accurate match points. Local activity could be measured by using the local busyness texture measure (t_B). This approach could be very effectively used in a system designed to approximately register images with the measurement of a few match points.

Once a match has been undertaken, information generated in the matching process can be used to further refine the estimate of match point accuracy. The value of the correlation parameter at the correlation peak gives a measure of texture similarity at the match point. A threshold can be set to require a certain level of texture similarity for a good match. The accuracy of match can be evaluated using the sharpness of the correlation peak. In general, the sharper the correlation peak, the more accurate the match.

6.0 SUMMARY

To solve the general matching problem of dissimilar images, a texture matching approach has been taken. It is assumed that from one image of a scene to another there are elements of texture which will remain invariant. This approach can potentially be used in a wide range of image matching applications. These range from the approximate registration problem where a few match points are required to the problem of determining thousands of match points. The key to this approach is the development of specific conversion algorithms for the pixels of each system of interest, so that the resultant pixel values are effectively system independent. The actual pixel replacement algorithms will have to be empirically tailored for each specific collection system.

To complement this completely automatic approach, a semi-automatic approach is presented which can be used for the approximate registration problem. This approach, computer assisted manual image registration (CAMIR), provides computer and special-purpose hardware assistance to the operator who is required to measure matching points from each of the images. This system could be readily constructed given the current state-of-the-art of digital hardware.

The first step in the match problem is to apply geometric corrections using the initial knowledge of the taking parameters. The inclusion of geometric effects requires three steps:

a) A grid of points is defined on the ground at positions at which it is desired to know the intensity values,

b) These grid points are mapped into the projection space and their corresponding positions in projection space are determined.

c) Pixel intensity values at these positions in projection space are determined using interpolation and the existing intensity measurements.

The geometrically corrected image is formed from these interpolated pixel values. After geometric correction, the pair of images to be matched will usually differ by a translational shift. This misregistration is due to residual uncertainties in the aim point determination. Rotational and scale differences between the images to be matched will normally be small. Application of geometric corrections greatly reduces the dimensions of the matching problem.

The CAMIR system makes use of the exceptional ability of a human operator to quickly solve complex pattern recognition problems. Computer and special high-speed digital hardware are combined in a manner to permit an operator to solve the registration problem with maximum efficiency. The operator has a CRT display to view each of the two dissimilar images. His problem is to find and measure three or more match points from the extremes of the images. The hardware permits the operator to have complete freedom to explore and examine the images. He can translate, rotate, and magnify or demagnify the sub-image being viewed. The ability to control image sharpness and noise can be provided under the real-time control of the operator. This is provided

by adjusting spatial response with a modulation transfer function compensator (MTFC). Similarly, the ability to modify scene intensities can be provided under the real-time control of the operator. This dynamic range adjustment (DRA) maps intensity levels in any desired manner.

To provide for maximum efficiency in registering images, geometric corrections are always applied before the images are displayed. To be able to apply corrections to a variety of taking systems, the transformations from object space to projection space are expressed in a general polynomial form. With this approach software can cast the correction problem for any system into a uniform form. Special purpose hardware can then solve the problem in real time. Once the three corresponding points have been measured by the operator, the images are brought together into registration. This is accomplished by adjusting the parameters of the distortion polynomials.

This system could be used as an integral part of a variety of systems designed to attack map making problems with digital techniques. These could include map updating systems or systems designed to build up contour data base information. This system can be built using currently available digital components. For E-Systems to build such a system, no research or development effort would be required. The only requirement to build the image manipulator would be engineering effort.

The texture matching approach to image matching has the advantage of being completely automatic. The first step in the process is to correct for all known geometric distortions. After geometric correction, the two images to be matched will normally differ in registration by a translation shift. In the next step, each of the images to be matched is transformed into a texture measure space. Each pixel value is replaced by some local measure of the surround (texture). It is not always necessary to replace each and every pixel intensity value by a texture measure, and the frequency of replacement that is used is referred to as the pixel replacement resolution. The match point is determined by correlating in the texture measure space. A reference window is chosen from one image and a larger search area is chosen from the other image. The normalized cross correlation product is used and the match point is determined by the position of the correlation peak. A threshold can be set using the value of the correlation parameter at the peak to determine if a good match has been found. This process can be repeated across the two images to obtain match points as frequently as desired. This technique is generalized to permit the use of multiple texture measures.

Local spectral analysis is an extremely effective tool that can be used to develop appropriate texture measures. The local spectral analysis can be performed with the use of the two dimensional discrete Fourier transform. This transform has the advantage of having a fast algorithm (the FFT) available for

its implementation. Texture measures can be defined by grouping together similar spectral components using the general prescription:

$$t = \sqrt{\sum_{u,v} \left| X_{u,v} \right|^2} \qquad (6\text{-}1)$$

Here $X_{u,v}$ represents th complex spectral components. Spectral components can be grouped according to their radial frequency ($\sqrt{u^2+v^2}$). A variety of possible texture measures are given for the four-by-four and eight-by-eight neighborhood sizes. The size of the neighborhood defines the local area used in determining the texture measures.

If it is desired to use a single measure of texture, the local busyness measure should be extremely effective. This is the texture measure which is used to demonstrate the texture matching approach. It is defined as the square root of the AC energy present in the pixel neighborhood.

To use texture matching, the algorithm must be empirically tailored to the combination of collection systems that are used. The parameters which must be experimentally determined include:

a) the selection of the texture measure or measures to be used,

b) the size of the reference window and search area to be used,

c) the size of the neighborhood used to determine the texture measure,

d) and the pixel replacement resolution to be used.

There are a variety of special problems or effects which can interfere with the matching process. A number of these effects are discussed and methods of dealing with these problems are presented. These problems include data noise characteristics, different detector spatial acceptances, drastically different intensity distributions, and coherency effects found in dealing with imaging radars.


DAT
FILM